



# Über die Charakteristika der Leistung in Konzentrationstests

Prozesskomponenten, Mechanismen und Übungseffekte

Dissertation zur Erlangung  
des Doktorgrades der Naturwissenschaften  
(Dr. rer. nat.)

dem Fachbereich Psychologie  
der Philipps-Universität Marburg vorgelegt von

Iris Blotenberg  
geboren in Rostock

Marburg (Lahn), im Februar 2019

Vom Fachbereich Psychologie der Philipps-Universität Marburg (Hochschulkennziffer 1180)

als Dissertation angenommen am: 11.04.2019

Datum der Disputation: 11.04.2019

Erstgutachter und Betreuer: Prof. Dr. Lothar Schmidt-Atzert

Zweitgutachter: Prof. Dr. Stefan Krumm

Mitglieder der Prüfungskommission:

Prof. Dr. Lothar Schmidt-Atzert

Prof. Dr. Stefan Krumm

Prof. Dr. Hanna Christiansen

Prof. Dr. Alexander Schütz

## **Danksagung**

Zuallererst gilt mein großer Dank Prof. Dr. Lothar Schmidt-Atzert, der mir in den letzten Jahren stets mit Rat und Tat zur Seite stand und mit mir in vielen Mittags- und Kaffeepausen über Gott und die Welt, natürlich besonders die Diagnostik im Allgemeinen und die Konzentrationsfähigkeit im Speziellen, plauderte. An diesen Kaffeerunden und regen Diskussionen beteiligten sich auch Dr. Karin Funsch und Dr. Beatriz Arias Martín, die mich ebenfalls an ihrem großen Schatz an Wissen und ihrer praktischen Erfahrung in der Diagnostik teilhaben ließen und mich über die Jahre sehr unterstützten. Außerdem möchte ich Prof. Dr. Stefan Krumm sehr für die Begutachtung dieser Arbeit danken. Mein weiterer Dank gilt Prof. Dr. Hanna Christiansen und Prof. Dr. Alexander Schütz, die sich für die Prüfungskommission zur Verfügung stellen. Ein großes Dankeschön gilt Prof. Dr. Sarah Teige-Mocigemba, die mich so herzlich in ihre neu gegründete Arbeitseinheit aufnahm und Zahra Khosrowtaj, Paul Bacher und Lea Nahon für eine kurze, aber sehr schöne gemeinsame Zeit des Doktorandendaseins. Schließlich gilt mein Dank meiner Familie, meinen Eltern Birga und Günther, meiner Oma Eva und ganz besonders meiner Schwester Jana und meinem Freund Jasper – danke, dass ihr immer für mich da seid.

## Inhaltsverzeichnis

<b>Zusammenfassung</b> .....	1
<b>Abstract</b> .....	4
<b>1. Einführung</b> .....	7
<b>2. Über die Konzentrationsfähigkeit</b> .....	11
2.1 Definitionsansätze und Modelle .....	11
2.2 Fazit .....	14
<b>3. Über die Messung der Konzentrationsfähigkeit</b> .....	15
3.1 Arten von Konzentrationstests.....	15
3.2 Charakteristika von Konzentrationstests .....	18
3.3 Fazit .....	20
<b>4. Korrelative Validitätsbelege für Konzentrationstests</b> .....	21
4.1 Konstruktvalidität – Konvergente Validität .....	21
4.2 Konstruktvalidität – Diskriminante Validität .....	22
4.3 Kriteriumsvalidität.....	23
4.4 Ein Schwachpunkt: Übungseffekte in Konzentrationstests.....	24
4.5 Fazit .....	26
<b>5. Experimentelle Validitätsbelege für Konzentrationstests</b> .....	27
5.1 Experimentelle Studien zur Rolle des Präsentationsmodus .....	27
5.2 Experimentelle Studien zur Koordination von Teilhandlungen .....	28
5.3 Fazit .....	29
<b>6. Zielsetzung des Dissertationsvorhabens</b> .....	31
<b>7. Empirischer Teil</b> .....	35
7.1 Beitrag 1: Auf dem Weg zu einem Prozessmodell für Konzentrationstests.....	35
7.2 Beitrag 2: Über die Charakteristika der Konzentrationsleistung.....	40
7.3 Beitrag 3: Zu den Ursachen des Übungseffekts in Konzentrationstests.....	46
<b>8. Gesamtdiskussion</b> .....	50

8.1 Komponenten der Leistung in Konzentrationstests.....	51
8.2 Der Übungseffekt in Konzentrationstests.....	59
8.3 Limitationen und Stärken .....	64
8.4 Fazit und Ausblick.....	65
<b>9. Literaturverzeichnis.....</b>	<b>68</b>
Anhang A: Manuskript zum Beitrag 1 .....	76
Anhang B: Manuskript zum Beitrag 2 .....	102
Anhang C: Manuskript zum Beitrag 3 .....	127
Curriculum Vitae.....	160
Eidesstattliche Erklärung.....	164

## Zusammenfassung

Tests zur Messung der Konzentrationsfähigkeit sind seit über einhundert Jahren im Einsatz und finden in den verschiedensten psychologischen Feldern – von der Verkehrs- über die Wirtschafts- bis zur Neuropsychologie – Anwendung (Schmidt-Atzert & Amelang, 2012). Typische Konzentrationstests präsentieren eine Reihe von homogenen Stimuli, die kontinuierlich nach einer einfachen Regel zu bearbeiten sind, bis die Zeit vorüber ist (Westhoff & Hagemeister, 2005). Für diese Tests liegen diverse korrelative Validitätsbelege vor, die für konvergente (in Form eines gemeinsamen Konzentrationsfaktors) und diskriminante Validitäten (im Sinne einer gelungenen Abgrenzung von höheren kognitiven Fähigkeiten) dieser Tests sprechen (z.B. Krumm et al., 2009; Schmidt-Atzert, Bühner, & Enders, 2006). Allerdings lassen sich einige Fragestellungen mittels korrelativer Studien nur schwer adressieren, hier erwiesen sich Ansätze der experimentellen Testvalidierung als fruchtbar, um Schlüsselannahmen über Konzentrationstests zu prüfen (z. B. Krumm, Schmidt-Atzert, Bracht, & Ochs, 2011; Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, Schmidt, Zenses, & Stenzel, 2012). Nichtsdestotrotz sind bislang wichtige Fragen offen geblieben: So gibt es noch kein Prozessmodell für das Zustandekommen der Leistung in Konzentrationstests und die relativ großen Übungseffekte in diesen Tests sind zwar gut belegt, aber kaum verstanden. Das Ziel der vorliegenden Arbeit lag darin, die Prozesskomponenten bei der Bearbeitung von Konzentrationstests und die relativ großen Übungseffekte in diesen kognitiven Tests anhand dreier Beiträge zu beleuchten.

Auf der Basis theoretischer und empirischer Arbeiten wird im *ersten Beitrag* ein generisches Prozessmodell der Leistung in Konzentrationstests vorgeschlagen. Demnach sollte die Leistung in einem Konzentrationstest davon abhängen, wie schnell eine Testperson 1) ein Item wahrnimmt, 2) eine simple mentale Operation durchführt, um ein Item zu lösen, 3) eine motorische Reaktion ausführt, um die Lösung zu indizieren, 4) selbstgesteuert zum nächsten Item wechselt. In zwei Studien ( $N_{\text{Studie 1}} = 103$ ,  $N_{\text{Studie 2}} = 100$ ) konnte gezeigt werden, dass die Geschwindigkeit in den einzelnen Prozesskomponenten die Leistung in drei verschiedenen Konzentrationstests maßgeblich vorhersagte (55 – 74 % Varianzaufklärung). Dabei waren besonders die Wahrnehmungsgeschwindigkeit und die Geschwindigkeit der Lösungsprozesse starke Prädiktoren für die Leistung im Test, während sich ein Trend für einen kleineren Einfluss der motorischen Geschwindigkeit zeigte. In Bezug auf die vierte postulierte Komponente, den selbstgesteuerten Itemwechsel, konnte zwar demonstriert

werden, dass es für die Leistung einen großen Unterschied machte, ob kontinuierlich gearbeitet und selbstgesteuert zwischen den Items gewechselt werden musste oder ob kurze Pausen zwischen den Items vorgesehen waren: Wenn die Testpersonen sukzessiv und ohne vorgesehene Pausen arbeiten mussten, reagierten sie langsamer und machten mehr Fehler. Allerdings konnten keine stabilen interindividuellen Unterschiede in der Geschwindigkeit des Itemwechsels erfasst werden und es zeigten sich keine Zusammenhänge mit der Leistung in Konzentrationstests. Insgesamt konnten im ersten Beitrag wichtige Erkenntnisse über die Prozesskomponenten der Leistung in Konzentrationstests gewonnen werden. Es blieb aber die Frage offen, wie sich die kontinuierliche Bearbeitung von Items auf der Prozessebene auswirkt. Die Anforderung zum kontinuierlichen Arbeiten, das hatten frühere Studien bereits gezeigt, hat einen starken Einfluss auf die Validität dieser Tests (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, et al., 2012).

Der *zweite Beitrag* befasst sich mit einer möglichen Erweiterung des Prozessmodells um weitere Komponenten. Dabei wird angenommen, dass die Darbietung vieler Stimuli, die in Konzentrationstests üblich ist, zum einen die Fokussierung auf den aktuell relevanten Stimulus erforderlich machen sollte, zum anderen aber auch die Vorschau auf nachfolgende Stimuli erlauben und zur Vorverarbeitung dieser nachfolgenden Stimuli genutzt werden könnte. Für diese Studie ( $N = 100$ ) wurde ein modifizierter d2 erstellt und der Präsentationsmodus so manipuliert, dass die Anforderung an die Fokussierung und die Möglichkeit zur Vorschau auf nachfolgende Stimuli systematisch variiert wurden. In Bezug auf die Fokussierungsleistung konnte kein Effekt gefunden werden. Dabei waren die gewählten ablenkenden Stimuli womöglich zu schwach. Allerdings zeigte sich, dass die Testpersonen die Vorschau auf nachfolgende Stimuli nutzten und diese vorverarbeiteten, sofern sie die Möglichkeit dazu bekamen. Interindividuelle Unterschiede darin, wie gut ihnen die Vorbereitung auf die nachfolgenden Stimuli gelang, waren reliabel und hingen mit der Leistung in verschiedenen Konzentrationstests zusammen. Somit war es im zweiten Beitrag gelungen, eine neue Komponente der Leistung in Konzentrationstests aufzudecken: Die Vorverarbeitung nachfolgender Stimuli. Dabei sind durchaus Zusammenhänge mit einigen alltäglichen Konzentrationsaufgaben denkbar, die ebenfalls von der proaktiven Verarbeitung von Umgebungsreizen profitieren, wie das Fahren eines Autos oder die Korrektur eines Textes. Es bleibt dennoch zu prüfen, inwieweit es sich bei dieser Komponente um einen inherenten Bestandteil der Konzentrationsfähigkeit oder um einen möglicherweise nicht intendierten Nebeneffekt der Testgestaltung handelt.

Der *dritte Beitrag* der vorliegenden Arbeit widmet sich den Übungseffekten in Konzentrationstests. Relativ große Übungseffekte zwischen einem Drittel und bis zu einer Standardabweichung nach einmaliger Testwiederholung sind ein gut dokumentiertes Phänomen in diesen Tests (z. B. Bühner, Ziegler, Bohnes, & Lauterbach, 2006; Scharfen, Peters, & Holling, 2018; Westhoff & Dewald, 1990). Es ist aber noch kaum verstanden, wie diese Übungseffekte zustande kommen (Büttner & Schmidt-Atzert, 2004; Scharfen, Blum, & Holling, 2018). Daher wird in dieser Untersuchung ( $N = 100$ ) das Prozessmodell der Konzentrationsleistung zugrunde gelegt und eine Reihe kognitiver Aufgaben zweimal dargeboten, um zu untersuchen, inwieweit die einzelnen Prozesskomponenten von der Übung profitieren. Es zeigte sich, dass sich die Übung an mehreren Stellen im Prozess auszuwirken schien, in kleinerem Maße auf Wahrnehmungs- und Motorikprozesse, in größerem Maße auf die simple mentale Operation zur Lösung des Items und wahrscheinlich auch auf die Koordination der Prozesskomponenten. Weder der selbstgesteuerte Wechsel zwischen den Items noch die Vorbereitung auf nachfolgende Items wurden mit der Testwiederholung effizienter. Im dritten Beitrag war es somit gelungen, auf der Prozessebene einen Einblick in die Entstehung der Übungseffekte in Konzentrationstests zu gewinnen. Es zeigte sich, dass die Auswirkungen der Übung in Konzentrationstests komplex sind, was es schwierig macht, sie durch Anpassungen des Testmaterials zu vermindern oder die Geübtheit versus Ungeübtheit einer Testperson erkennbar zu machen.

Zusammengenommen ermöglichen die vorgestellten Beiträge tiefere Einblicke in die Prozesskomponenten, die bei der Bearbeitung von Konzentrationstests eine Rolle spielen und damit ein besseres Verständnis, was diese Tests messen. Dabei ist eine Vielzahl kognitiver Prozesse an diesen so simpel anmutenden Aufgaben beteiligt. Die Ergebnisse dieser Arbeit werfen auch neue Fragen auf, beispielsweise bleibt offen, wie sich die kontinuierliche Bearbeitung von Stimuli auf der Prozessebene im Detail auswirkt, inwieweit die Fähigkeit zur Vorverarbeitung nachfolgender Stimuli in Konzentrationstests mit der proaktiven Verarbeitung in alltäglichen Konzentrationsaufgaben korrespondiert und wie sich die Übung in Konzentrationstests auf die Validität dieser Tests auswirkt. Alles in allem kann die psychologische Diagnostik sehr von experimentellen Validierungsansätzen profitieren. Für die Zukunft wäre es vielversprechend, diese zusätzlich um Ansätze der mathematischen Modellierung kognitiver Prozesse zu ergänzen.



## Abstract

Sustained attention tests have been in use for more than a hundred years and they are applied in various psychological disciplines from traffic to business to neuropsychology (Schmidt-Atzert & Amelang, 2012). Typically, these tests consist of many homogenous and simultaneously presented stimuli and the test-taker is required to continuously apply a simple rule to them (like finding targets or assessing whether two single-digit numbers add up to a third) until the test is over (Westhoff & Hagemester, 2005). A large body of correlational validity evidence supports sustained attention tests' convergent (there is evidence for a common factor; e.g. Schmidt-Atzert et al., 2006) and discriminant validity (correlations with higher cognitive abilities are low to moderate; e.g. Krumm et al., 2009). However, some research questions are difficult to address using only correlational approaches to validity. For example, until today, there is no generic process model of performance in sustained attention tests and the large practice effects in these tests are well confirmed, but poorly understood. Therefore, a more in-depth analysis of the processes involved in sustained attention tests requires an experimental approach. Along these lines, the aim of the present thesis was to investigate the processes involved in sustained attention tests and the locus of the practice effect in these tests.

In the first contribution of the present thesis, based on theoretical considerations and earlier research, a process model of sustained attention tests was proposed. According to the model, performance in these tests should depend on: How fast the test-taker 1) perceives an item, 2) performs a simple mental operation to solve an item, 3) responds to indicate the answer to an item and 4) deliberately shifts towards the next item. In two studies ( $N_{\text{Study 1}} = 103$ ,  $N_{\text{Study 2}} = 100$ ), three sub-components jointly predicted 55 to 74 % of the variance in three sustained attention tests. Perceptual and item solving processes showed to be strong predictors of performance, while there was a consistent trend towards a small influence of motor speed on test performance. Regarding item shifting, it was confirmed that the requirement to constantly stay on task and deliberately shift between items (the so-called self-paced mode) affected performance compared to a force-paced mode (sometimes also called computer-paced mode) which involved short intervals between successive stimuli. However, interindividual differences in item shifting were low in reliability and unrelated to performance in sustained attention tests. Thus, it remains to be elucidated how the requirement to continuously work on and deliberately shift between items impacts information processing in sustained attention tests.

In the second contribution of the present thesis, it was examined whether the process model had to be extended when considering the characteristic presentation mode of sustained attention tests. Typically, in these tests, many stimuli are presented simultaneously and should therefore require the test-takers to focus on the currently relevant stimulus but could also enable them to preprocess upcoming stimuli. In order to assess the role of focusing and preprocessing in sustained attention tests, a modified version of the d2 test of sustained attention was created and its stimulus arrangement was manipulated. There was no effect of focusing in the present data, which might be due to the use of relatively simple distractors. However, it was shown that there was a large preview benefit (with regard to speed and error rates) in the conditions that allowed a preprocessing of upcoming stimuli. Moreover, interindividual differences in the extent to which preprocessing took place were reliable and substantially correlated with performance in three different sustained attention tests. Thus, a new component of performance in sustained attention tests was revealed: The preprocessing of upcoming stimuli. This component of test performance could very likely be linked to everyday sustained attention tasks that also benefit from a proactive scanning of potentially relevant information in the environment, like driving a car or proofreading a text. However, this potential link is yet to be investigated before we can draw conclusions about whether preprocessing constitutes an inherent part of sustained attention or rather an ancillary effect of the way sustained attention tests are designed.

The subject of the third contribution of the present thesis were the large practice effects in sustained attention tests that have frequently been reported and range between one-third and one standard deviation (z. B. Bühner et al., 2006; Scharfen, Peters, et al., 2018; Westhoff & Dewald, 1990). Based on the previously proposed process model, it was investigated which of the sub-components of sustained attention tests benefitted from practice and to which extent. Therefore, in this study ( $N_{\text{Study}} = 100$ ), several cognitive tasks were administered twice to examine the effects of practice on the sub-components. It was shown that practice affected several sub-components, among them perceptual and motor processes but especially item solving processes and potentially also the coordination of the sub-components. However, neither the deliberate shifting between items nor the preprocessing of upcoming items became more efficient through practice. Altogether, the effects of practice on the sub-components of sustained attention tests seem to be rather complex and therefore, taking measures in order to reduce the practice effect in these tests or to recognize pre-exposure to these tests appears challenging.

To sum up, using experimental test validation approaches, the present thesis allows a more in-depth insight into the processes involved as participants work through sustained attention tests and therefore, allows a better understanding of what these tests measure. It has again been shown that performance in these seemingly simple tests relies on several cognitive processes. However, the results of the current studies also raise new questions: For example, it is yet to be investigated how the continuous self-paced processing of stimuli impacts information processing in these tests, whether preprocessing corresponds to everyday sustained attention tasks and finally, how practice affects the validity of sustained attention tests. Altogether, the field of psychological assessment can strongly benefit from experimental approaches to validity. In future studies, it might be promising to also apply mathematical models of cognitive processes to address some of the open questions.

## 1. Einführung

*„If there is a single challenge that characterizes all of psychology’s diverse subfields, that challenge is assessment.“ (Bornstein, 2011, S. 532)*

Das Herzstück der Güte eines psychologischen Testverfahrens ist seine Validität, also die Frage, ob der Test misst, was er zu messen vorgibt. Seit über fünfzig Jahren liegt der Standard, um diese Validitätsbelege zu erbringen, in korrelativen Studien zur Konstrukt- und Kriteriumsvalidität (Bornstein, 2011; Borsboom, Mellenbergh, & Van Heerden, 2004). In zahlreichen Studien gehen Testentwickler entlang nomologischer Netzwerke der Frage nach, ob und wie hoch ein Test mit anderen Verfahren korreliert, die etwas Ähnliches messen sollen, wie der Zusammenhang mit Tests ist, die etwas Anderes messen sollen, und wie das Testergebnis mit Außenkriterien außerhalb des Labors zusammenhängt.

Seit einiger Zeit wird der Wunsch lauter, diese Validitätsbelege zu ergänzen und zu untersuchen, welche Prozesse und Mechanismen bei der Testbearbeitung eine Rolle spielen (Bornstein, 2011; Borsboom et al., 2004; Embretson, 1983; Krumm, Hüffmeier, & Lievens, 2017). Verstünde man besser, wie ein Testergebnis zustande kommt, so die Überlegung, so würde man auch ein tieferes Verständnis dafür entwickeln, was ein Test eigentlich misst und warum er mit anderen Testverfahren in welchem Ausmaß korreliert. Im Fokus steht dabei die Frage, ob das zu messende Konstrukt tatsächlich verursachend für das Testergebnis ist. Tatsächlich fordern auch die neuesten *Standards for Educational and Psychological Assessment*, welche gemeinsam von der American Educational Research Association, der American Psychological Association und dem National Council on Measurement in Education herausgegeben werden, dass Validitätsbelege über bloße Angaben zur Struktur des Tests und seiner Beziehung zu anderen Maßen hinausgehen sollten (AERA, APA, & NCME, 2014). Es sollten daher auch Untersuchungen zum Antwortprozess hinzugezogen und geprüft werden, ob dieser den theoretischen Vorstellungen entspricht, das heißt, ob die beteiligten Prozesse konstruktkonform sind und somit nahelegen, dass der Test tatsächlich misst, was er zu messen vorgibt.

Dieser Gedanke soll anhand eines kleinen Beispiels näher ausgeführt werden: Zur Messung der fluiden Intelligenz nach Cattell (1963) werden häufig Matrizenaufgaben eingesetzt, in denen mehrere Figuren dargeboten werden, die sich zum Beispiel in Form, Muster,

Komplexität, Größe, Anzahl und relativer Position von Teilen der Figur unterscheiden. Es fehlt dabei jeweils eine Figur, diese ist unter den vorgegeben Antwortalternativen auszuwählen. Damit dies gelingen kann, sollte der Itemstamm von der Testperson auf Gesetzmäßigkeiten geprüft und auf dieser Basis die richtige Lösung unter den Antwortalternativen ausgewählt werden. Die Fähigkeit zum Erkennen von Gesetzmäßigkeiten und zum schlussfolgernden Denken sollte also dazu führen, dass die Testperson das Item richtig lösen kann. So würde das Ergebnis im Test die Fähigkeit der Testperson zum schlussfolgernden Denken widerspiegeln. Nun konnten Mittring und Rost (2008) allerdings zeigen, dass sich einige Items verschiedener Matrizentests auch durch ein Ausschlussverfahren der Distraktoren lösen lassen (siehe auch White & Zammarelli, 1981). Dafür ist nicht entscheidend, dass die Gesetzmäßigkeiten im Itemstamm von der Testperson richtig erkannt werden, stattdessen werden die Distraktoren nach relativ simplen Regeln erkannt und ausgeschlossen. Das bedeutet, dass hier nicht unbedingt die Fähigkeit zum schlussfolgernden Denken, die gemessen werden soll, zur richtigen Lösung (und später zu einer hohen Punktzahl im Test) geführt haben muss, sondern dass eine alternative Strategie ebenfalls zum richtigen Ergebnis führt – eine Bedrohung für die Validität dieser Tests. In der Folge wurden Matrizentests mit einem offenen Antwortformat entwickelt, die eine solche alternative Strategie nicht erlauben (z.B. Becker, Preckel, Karbach, & Raffel, 2014).

Die besondere Bedeutung der Validität als wichtigstes Gütekriterium im Allgemeinen und der Prozessvalidierung im Speziellen, zeigt sich auch darin, dass diese Themen im ersten Kapitel der *Standards for Educational and Psychological Assessment* behandelt werden (AERA et al., 2014). Eine Möglichkeit, um diese Prozesse zu untersuchen, ist die experimentelle Testvalidierung – dabei bedienen sich Forscher experimenteller Methoden, um die Prozesse zu analysieren, die bei der Bearbeitung eines Tests beteiligt sind, beispielsweise durch die systematische Manipulation von wichtigen Testmerkmalen (Bornstein, 2011; Borsboom et al., 2004; Embretson, 1983; Krumm et al., 2017). Ein solches Vorgehen erlaubt nicht nur ein vertieftes Verständnis des jeweiligen Tests und der Fähigkeiten und Prozesse, die während der Bearbeitung gefordert sind. Es ermöglicht auch die nähere Untersuchung von Phänomenen, die bei psychologischen Tests beobachtet werden, aber bislang kaum verstanden sind. Beispielsweise sind die meisten kognitiven Leistungstests mit Übungseffekten behaftet, aber es ist bisher wenig untersucht, an welcher Stelle im Antwortprozess diese im Detail auftreten, also welche Prozesse oder Mechanismen eigentlich so stark von der Übung profitieren (siehe auch Scharfen, Peters, et al., 2018, für eine aktuelle Metaanalyse).

In diesem Sinne soll in der vorliegenden Arbeit untersucht werden, welche Prozesse und Mechanismen eine wesentliche Rolle für die Leistung in Konzentrationstests spielen. Die Bedeutung dieser Tests ist hoch, gehören sie doch zu den am häufigsten eingesetzten psychologischen Testverfahren (Schmidt-Atzert & Amelang, 2012; Westhoff & Hagemeyer, 2005). Von der Neuropsychologie über die Verkehrspsychologie, von der pädagogischen Psychologie bis hin zur psychologischen Eignungsdiagnostik – die Konzentrationsleistung ist in nahezu allen psychologischen Anwendungsfeldern von Interesse. Dennoch haben diese Tests und ihr zugrundeliegendes Konstrukt im Vergleich zu anderen kognitiven Fähigkeiten, wie der Intelligenz- oder der Arbeitsgedächtnisleistung, in der Forschung vergleichsweise wenig Beachtung erfahren (Beckmann & Strang, 1993; Schweizer & Moosbrugger, 2004). Für die letztgenannten Konstrukte existiert eine überwältigende Zahl an wissenschaftlichen Abhandlungen und Veröffentlichungen sowie viel beforschte Struktur- und Prozessmodelle (für einen Überblicksartikel siehe McGrew, 2009). Solche Modelle sind für die Konzentrationsfähigkeit und die Leistung in Konzentrationstests noch nicht in vergleichbarem Maße vorhanden (Schmidt-Atzert & Amelang, 2012). So gibt es bisher keinen verbindlichen konzeptionellen Rahmen und kein etabliertes Prozessmodell (Schmidt-Atzert, Büttner, & Bühner, 2004; Schmidt-Atzert et al., 2006). Außerdem wurden weitere potentiell relevante Mechanismen, die für die Leistung in diesen Tests eine Rolle spielen könnten, bisher kaum untersucht. Dazu zählen beispielsweise die Fokussierung auf den aktuell relevanten Stimulus oder die proaktive Vorverarbeitung nachfolgender Stimuli. Auch die Ursachen des häufig berichteten, relativ großen Übungseffekts in Konzentrationstests sind weitgehend ungeklärt (z.B. Scharfen, Blum, et al., 2018; Schmidt-Atzert et al., 2004). Das Ziel der vorliegenden Arbeit besteht darin, sich diesen Fragen zu nähern und Ansätze der experimentellen Testvalidierung zu nutzen, um besser zu verstehen, wie die Leistung in Konzentrationstests zustande kommt und letztlich, ein tieferes Verständnis dafür zu entwickeln, welche Fähigkeiten diese Tests messen. Drei wesentliche Leitfragen werden behandelt:

- 1) Auf dem Weg zu einem Prozessmodell: Welche Prozesskomponenten leisten einen Beitrag zur Leistung in Konzentrationstests?*
- 2) Welche weiteren Mechanismen und Fähigkeiten spielen für die erfolgreiche Bearbeitung von Konzentrationstests eine Rolle?*
- 3) Welche der Prozesskomponenten profitieren in welchem Ausmaß von der Übung?*

Mit der vorliegenden Arbeit sollen weitere Schritte auf dem Weg zur Beantwortung dieser Fragen gegangen werden. Bevor wir uns diesen Fragen widmen, werden einige theoretische Überlegungen zur Konzentrationsfähigkeit, zu Konzentrationstests, sowie wichtige Forschungsergebnisse ausgewählter Studien vorgestellt.

## 2. Über die Konzentrationsfähigkeit

Einen Fachtext lesen, eine Prüfung ablegen, ein Auto durch den dichten Verkehr steuern, Rechnungen überprüfen, einen Text korrigieren – vielerlei alltägliche Aufgaben verlangen uns die Fähigkeit ab, uns über längere Zeiträume kontinuierlich einer Aufgabe zuzuwenden. Diese Fähigkeit wird als Konzentration oder auch als die Fähigkeit zum konzentrierten Arbeiten bezeichnet und gilt als Voraussetzung für höhere kognitive Fähigkeiten (Lezak, 1995; Schweizer, Moosbrugger, & Goldhammer, 2005; Westhoff, 1995). Ist sie gemindert, so leiden schulische und berufliche Leistungen, zudem äußern sich einige neurologische und psychische Störungen unter anderem durch eine herabgesetzte Konzentrationsfähigkeit (Schmidt-Atzert & Amelang, 2012). Bei dieser kognitiven Fähigkeit handelt es sich demnach gewissermaßen um eine Schlüsselfähigkeit, der diese Arbeit gewidmet wird.

### 2.1 Definitionsansätze und Modelle

Ein verbindlicher konzeptueller Rahmen für die Konzentrationsfähigkeit fehlt bislang (Goldhammer & Moosbrugger, 2006; Schmidt-Atzert et al., 2004, 2006). Das liegt sicherlich auch daran, dass die Abgrenzung dieses Konstrukts von verwandten kognitiven Fähigkeiten, wie der Aufmerksamkeit, schwerfällt. Einige Autoren ordnen die Konzentrationsfähigkeit in Modelle der Aufmerksamkeit ein und sehen in ihr ein Zusammenwirken verschiedener Aufmerksamkeitskomponenten, die im Sinne des supervisorischen Aufmerksamkeitssystems nach Shallice (1982) erhöhte Kontroll- und Koordinationsanforderungen stellen (Goldhammer & Moosbrugger, 2006). Andere Autoren trennen die beiden Fähigkeiten und argumentieren, dass sich die Aufmerksamkeit verstärkt auf Wahrnehmungsprozesse und die Selektion von Reizen bezieht, während sich die Konzentrationsleistung vor allem als Form des kontrollierten und koordinierten (Weiterver-) Arbeitens zeigt, die mit kognitiver Anstrengung einhergeht (Schmidt-Atzert et al., 2004; Westhoff, 1995). *Abbildung 1* veranschaulicht diese Annahme modellhaft. Auch im Verhältnis der Konzentrationsfähigkeit zu Intelligenzleistungen herrscht keine Einigkeit, so gilt sie manchen als Voraussetzung, anderen sogar als Basis der Intelligenz (Schweizer, 2005; Schweizer, Zimmermann, & Koch, 2000; Westhoff, 1995).



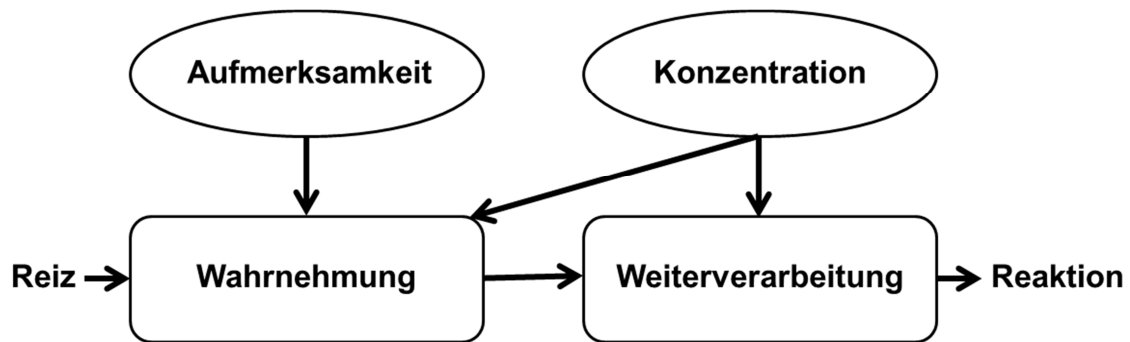


Abbildung 1. Das Verhältnis von Aufmerksamkeit und Konzentration. Dabei bezieht sich die Aufmerksamkeit auf die Wahrnehmung und die Konzentrationsleistung vorwiegend auf das Arbeiten (Schmidt-Atzert et al., 2004, S. 11).

Zwar gibt es bisher keine allgemein akzeptierte Definition der Konzentrationsfähigkeit, allerdings gibt es einige Vorschläge, die unterschiedliche Aspekte dieses Konstrukts beleuchten. Westhoff und Hagemeister (2005) greifen in ihrer Definition das bereits von Heinrich Düker (1957) genannte Koordinationskonzept auf und betonen darüber hinaus die Intentionalität und die Kontrollanforderungen, die beim konzentrierten Arbeiten nötig sind (siehe auch Berg & Imhof, 2001). Eine konzentrierte Person wendet sich demnach willentlich einer Aufgabe zu und koordiniert und kontrolliert die einzelnen Teilhandlungen, die zur Aufgabenerfüllung nötig sind.

*"Wir definieren also die Konzentration als die absichtsvolle, nicht automatisierte Koordination von Handlungsteilen und deren kontrollierte Ausführung."*  
(Westhoff & Hagemeister, 2005, S. 5)

Die Charakteristika der Konzentrationsleistung verdeutlicht Westhoff (1995) zusätzlich in seinem Akku-Modell der Konzentration. Darin wird die Konzentrationsleistung mit dem Akku einer Kamera verglichen, bei der eine Reihe von Teilhandlungen (z. B. das Blitzlicht, die Verengung oder Erweiterung der Linse, die Aufnahme, Projektion und Speicherung des Lichtbilds) energetisiert, koordiniert und kontrolliert werden müssen, damit eine Aufnahme entstehen kann. Zwar gibt es unterschiedlich starke Akkus, ihnen allen ist aber gemein, dass sie durch die kontinuierliche Nutzung an Energie verlieren, so dass sie von Zeit zu Zeit aufgeladen werden müssen. Schließlich ist der Akku in seiner Leistungsfähigkeit auch von Rahmenbedingungen, wie der Außentemperatur, abhängig. Übertragen auf die menschliche Konzentrationsfähigkeit bedeutet dies, dass diese nötig ist um die Teilhandlungen einer

Aufgabe über die Zeit zu energetisieren, zu koordinieren und zu kontrollieren, was mit Anstrengung verbunden ist (siehe auch Düker, 1957; Schmidt-Atzert et al., 2004). Menschen unterscheiden sich in ihrer Konzentrationsfähigkeit, allerdings leert sich auch der stärkste „Akku“ mit der Zeit, so dass jeder Mensch von Zeit zu Zeit Pausen benötigt, um sich im Anschluss wieder konzentrieren zu können. Schließlich ist die Konzentrationsfähigkeit auch von äußeren Faktoren, wie der Tageszeit, der Müdigkeit, der Einnahme bestimmter Substanzen (z.B. Kaffee) und vielen weiteren Faktoren abhängig. In der vorliegenden Arbeit geht es vorwiegend um die interindividuellen Unterschiede im Merkmal der Konzentrationsfähigkeit, wohl wissend, dass dieses Merkmal tagesformabhängigen Schwankungen unterliegen kann.

Ein weiteres wichtiges Charakteristikum der Konzentrationsleistung, nämlich ihr zeitlicher Aspekt, wird in einer Definition nach Schweizer (2005) hervorgehoben. Demnach liegt die Konzentrationsfähigkeit darin, sich kontinuierlich und über einen längeren Zeitraum auf eine Aufgabe zu fokussieren und dabei Störreize zu inhibieren. Im betreffenden Übersichtsartikel wird außerdem bekräftigt, dass diese Fähigkeit eine wesentliche Voraussetzung für die Bearbeitung komplexerer Aufgaben, also zum Beispiel von Intelligenzaufgaben, ist.

*„Die Fähigkeit, Verarbeitungsressourcen für eine längere Zeit (bis zu mehreren Minuten) auf eine bestimmte Aufgabe zu lenken und dabei andere Stimuli zu ignorieren, die auch Beachtung verlangen.“ Schweizer (2005, S. 46)*

In einer weiteren Definition der Konzentrationsfähigkeit widmen sich Schmidt-Atzert et al. (2004) einem wesentlichen Problem bei der Messung der Konzentrationsfähigkeit. Demnach ist der konzentrierte Zustand mit kognitiver Anstrengung verbunden (siehe auch Düker, 1957; Westhoff, 1995), die sich aber nur schwer operationalisieren lässt. Ein Phänomen, das mit dieser kognitiven Anstrengung einhergeht, lässt sich dagegen messen: So sollte sich die Fähigkeit zur Konzentration bei der Bearbeitung einer Aufgabe auch unter erschwerenden Randbedingungen (z.B. Lärm, sonstige Störreize, Müdigkeit) leistungssteigernd auswirken, was dazu führen sollte, dass schneller gearbeitet werden kann und weniger Fehler passieren. Dieser leistungssteigernde Aspekt der Konzentrationsfähigkeit lässt sich im Sinne einer Konzentrationsmessung operationalisieren.

*„Konzentration ist die Fähigkeit, unter Bedingungen schnell und genau zu arbeiten, die das Erbringen einer kognitiven Leistung normalerweise erschweren“. Schmidt-Atzert et al. (2004, S. 9)*

## **2.2 Fazit**

Zusammenfassend lässt sich feststellen, dass es bisher keine einheitliche Definition der Konzentrationsfähigkeit gibt (Goldhammer & Moosbrugger, 2006; Schmidt-Atzert et al., 2004, 2006), aber einige überzeugende Vorschläge, die verschiedene Aspekte dieser kognitiven Fähigkeit hervorheben und sich durchaus integrieren lassen. Als Synthese der vorgestellten Ansätze ließe sich Konzentration definieren als die Fähigkeit, über einen längeren Zeitraum kontinuierlich und fokussiert an einer Aufgabe zu arbeiten, dabei die Teilschritte dieser Aufgabe selbstgesteuert zu koordinieren und zu kontrollieren, sowie Störreizen und erschwerende Bedingungen zu trotzen, so dass die Aufgabe schnell und genau gelöst werden kann. Dabei geht konzentriertes Arbeiten mit einem hohen energetischen Aufwand und dementsprechend, mit mentaler Anstrengung einher (Düker, 1957; Schmidt-Atzert et al., 2004; Westhoff, 1995).

### 3. Über die Messung der Konzentrationsfähigkeit

Die Konzentrationsfähigkeit spielt für viele alltägliche und berufliche Aufgaben eine wichtige Rolle. Daher verwundert es nicht, dass Konzentrationstests sehr häufig eingesetzt werden (Schmidt-Atzert & Amelang, 2012) und schon, wenn man die relativ kurze Geschichte der wissenschaftlichen Psychologie bedenkt, früh entwickelt wurden. Wie bereits im vorherigen Abschnitt dargelegt, existiert bis heute keine einheitliche Definition und kein verbindliches Konzept der Konzentrationsfähigkeit (Goldhammer & Moosbrugger, 2006; Schmidt-Atzert et al., 2006). Daher legen unterschiedliche Autoren eigene Definitionen dieser kognitiven Fähigkeit zugrunde, was zu einer recht heterogenen Gruppe von Konzentrationstests geführt hat. Diese werden nach Bartenwerfer (1964) auch häufig unter dem Begriff der „allgemeinen Leistungstests“ zusammengefasst.

#### 3.1 Arten von Konzentrationstests

Einer der ersten standardmäßig eingesetzten Konzentrationstests stammt aus dem Jahre 1895 und geht auf den französischen Psychologieprofessor Benjamin Bienaimé Bourdon zurück (Bourdon, 1895). Dieser entwickelte diesen Test um die visuelle Diskriminationsfähigkeit von Schulkindern zu untersuchen und verwendete dafür verschiedene Testmaterialien, unter anderem Punkt-Muster und auch Buchstaben. Bei der Punkt-Muster Aufgabe werden eine Reihe von Punkt-Mustern, also beispielweise drei Punkte, die ein Dreieck formen oder vier Punkte, die wie ein Quadrat angeordnet sind, dargeboten. Die Aufgabe der Testperson besteht darin, unter den vielen dargebotenen Punktmustern diejenigen zu finden, die aus genau vier Punkten bestehen und von diesen so viele wie möglich in einer begrenzten Zeit wegzustreichen. Mit der Entwicklung dieses Tests begründete Bourdon die Gruppe der Durchstreichtests, deren moderne Vertreter, wie zum Beispiel der Konzentrationstest d2-R (Schmidt-Atzert & Brickenkamp, 2017) oder das Frankfurter Aufmerksamkeits-Inventar (FAIR, Moosbrugger & Oehlschlägel, 1996) sich noch immer einer großen Beliebtheit erfreuen. Ein weiterer früher Konzentrationstest ist der Pauli-Test (Arnold, 1975), bei dem sehr viele einstellige Zahlen untereinander auf einem Rechenbogen stehen und fortlaufend je zwei Zahlen zu addieren sind. Bis heute wurden noch eine Reihe weiterer numerischer Konzentrationstests entwickelt, darunter der Revisions-Test (Marschner, 1980), in dem viele einfache Gleichungen (simple Addition oder Subtraktion) dargeboten werden, die von der Testperson auf Richtigkeit zu prüfen und im Falle eines richtigen Ergebnisses abzhaken, im Falle eines falschen Ergebnisses wegzustreichen sind.

Eine weitere typische Aufgabengruppe unter den Konzentrationstests ist die der Sortiertests, bei denen Items anhand verschiedener Kriterien Kategorien zuzuordnen sind bis die Zeit abgelaufen ist. Der bekannteste Vertreter dieser Gruppe ist der Konzentrations-Verlaufs-Test (KVT) nach Abels (1974). Eine vierte Gruppe unter den Konzentrationstests ist die der Transformationsaufgaben. Zu dieser Gruppe gehört beispielsweise auch der Zahlen-Symbol-Test, der zwar als Test der Verarbeitungsgeschwindigkeit bezeichnet wird, aber alle Kriterien gängiger Konzentrationstests erfüllt (Krumm et al., 2011; Schmidt-Atzert et al., 2006, siehe auch Kapitel 4.2 „Diskriminante Validität“). In diesem Test wird der Testperson eine Liste mit Zahlen vorgelegt, denen jeweils ein Symbol zugeordnet ist. Darunter sind Reihen von Zahlen dargestellt unter denen sich leere Kästchen befinden. Die Aufgabe der Testperson besteht nun darin, die leeren Kästchen mit den passenden Symbolen zu füllen. In *Abbildung 2* ist das Aufgabenmaterial ausgewählter Konzentrationstests beispielhaft dargestellt.

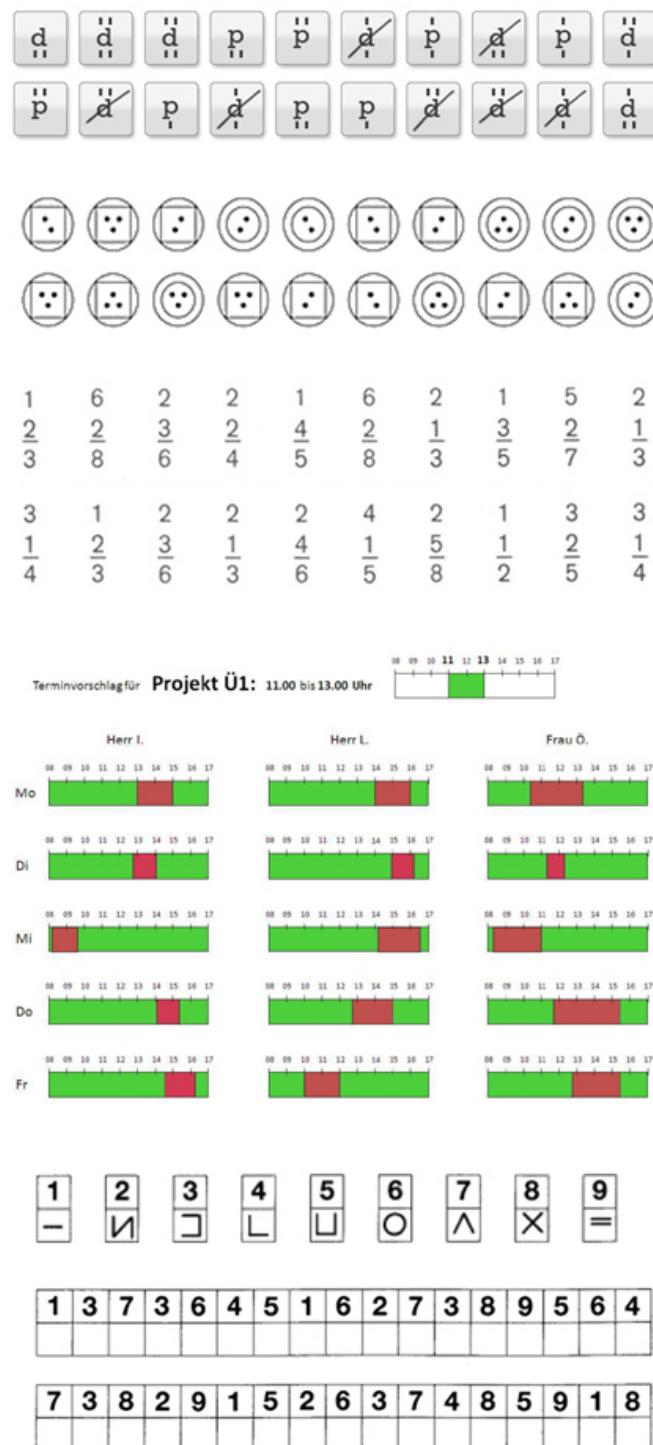


Abbildung 2. Beispielitems verschiedener Konzentrationstests. Von oben nach unten: Der Aufmerksamkeits- und Konzentrationstest d2-R (elektronische Fassung; Schmidt-Atzert & Brickenkamp, 2017) und das Frankfurter Aufmerksamkeits-Inventar 2 (FAIR-2; Moosbrugger & Oehlschlägel, 2011) stehen in der Tradition der Durchstreichkonzentrationstests. Der Revisions-Test (Marschner, 1980) ist ein Vertreter der Rechenkonzentrationstests und der Sortiertest (Bonnaire, 2014) dient als Beispiel für die Gruppe der Sortiertests. Der Zahlen-Symbol-Test (Jäger, Süß, & Beauducel, 1997) ist der wohl bekannteste Vertreter unter den Transformationstests.

### 3.2 Charakteristika von Konzentrationstests

Wie bereits im vorherigen Abschnitt beschrieben, ist die Gruppe der Konzentrationstests recht heterogen. Dennoch, und darauf wird später noch näher eingegangen, zeigen verschiedene Studien, dass diese Tests hoch miteinander korrelieren und auf einem gemeinsamen Faktor laden, also etwas Gemeinsames erfassen (Mirsky, Anthony, Duncan, Ahearn, & Kellam, 1991; Schmidt-Atzert et al., 2006). Es gibt tatsächlich einige grundlegende Gemeinsamkeiten dieser Tests, die unter anderem von Westhoff und Hagemeister (2005) benannt worden sind: So verwenden sie üblicherweise einfache, homogene Stimuli und stellen simple Aufgaben, für die eine einfach merkbare Regel anzuwenden ist. Der Grund hierfür ist, dass die Konzentrationsfähigkeit möglichst rein erfasst werden soll. Konzentration richtet sich allerdings immer auf eine konkrete Aufgabe und ist dementsprechend auch nur bei der Bearbeitung einer bestimmten Aufgabe erfassbar (Schmidt-Atzert et al., 2004, 2006). Für die Testentwicklung folgt daraus, dass Aufgaben verwendet werden sollten, die möglichst geringe Anforderungen an weitere Fähigkeiten und Fertigkeiten (z.B. Intelligenz, Problemlösefähigkeiten, Lesen, Rechnen, Merkfähigkeit etc.) stellen, damit die Konzentrationsmessung in möglichst geringem Maße mit diesen weiteren Fähigkeiten und Fertigkeiten konfundiert ist.

Da Testmaterial und Aufgaben so einfach sind, sollten die Testpersonen jedes Item problemlos lösen können, wenn sie nur ausreichend Zeit hätten. Den Testpersonen steht allerdings nur eine begrenzte Zeit zur Verfügung. Dieser Zeitdruck führt zu Leistungsunterschieden in der Anzahl der bearbeiteten Items und auch in der Anzahl der Fehler. Kennwerte von Konzentrationstests sind also typischerweise die Schnelligkeit – gemessen als die Anzahl der bearbeiteten Items – und die Genauigkeit – erfasst als die Anzahl der Fehler (Schmidt-Atzert et al., 2006; Westhoff & Hagemeister, 2005). In einigen Tests werden diese beide Kennwerte außerdem zur fehlerkorrigierten Geschwindigkeit verrechnet. Hintergrund für die Wahl dieser Kennwerte ist, dass die Konzentrationsfähigkeit als eine kognitive Fähigkeit betrachtet wird, die sich leistungssteigernd auswirkt. Das kennen wir auch aus dem Alltag: Unterlaufen uns Fehler (z.B. Tippfehler in einer Mail) oder brauchen wir ungewöhnlich lange für eine triviale Aufgabe, die uns normalerweise leicht fallen würde, erklären wir das oft mit mangelnder Konzentration. Das heißt, jemand, der konzentrierter ist, sollte schneller und exakter arbeiten als jemand, der die gleiche Aufgabe weniger konzentriert bearbeitet. Letzterer sollte langsamer vorankommen und dabei mehr Fehler machen. Auf diesen Zusammenhang beziehen sich auch Schmidt-Atzert et al. (2004, S. 9) in ihrer

Definition der Konzentrationsfähigkeit, die im vorigen Kapitel vorgestellt wurde. Danach zeigt sich die Konzentrationsfähigkeit im schnellen und genauen Arbeiten unter erschwerenden Bedingungen. Eine der erschwerenden Bedingungen im Konzentrationstest besteht demnach in der Zeitbegrenzung und dem damit verbundenen Zeitdruck, der Leistungsunterschiede in der Geschwindigkeit und auch in den Fehlern erzeugt.

Ein weiteres Merkmal von Konzentrationstests ist ihr charakteristischer Präsentationsmodus (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, et al., 2012; Schmidt-Atzert, Krumm, & Bühner, 2008; Westhoff & Hagemeister, 2005). Diesem kommt für die Messung der Konzentrationsfähigkeit eine ganz zentrale Rolle zu: Typischerweise werden viele Items gleichzeitig dargeboten und die Testperson ist gefordert, diese kontinuierlich zu bearbeiten bis die Testung vorbei ist. Das bedeutet, in diesem Präsentationsmodus sind ganz explizit keine Pausen vorgesehen, stattdessen sollen die Items für mehrere Minuten in selbstbestimmten Tempo bearbeitet werden. Bezogen auf energetische Aspekte der Konzentrationsleistung und das bereits erwähnte Akku-Modell der Konzentration nach Westhoff (1995) bedeutet das, dass es während des Tests und zwischen den Items keine Möglichkeiten gibt, den „Akku aufzuladen“, weil die einzelnen Teilhandlungen kontinuierlich energetisiert und koordiniert werden müssen. Für einen typischen Durchstreich-Test könnten diese Teilhandlungen beispielsweise in der Wahrnehmung der Stimuli, dem Abgleich mit dem Zielreiz, der Entscheidung, der motorischen Reaktion sowie dem Wechsel zum nächsten Item bestehen (siehe auch *Beitrag 1*, Blotenberg & Schmidt-Atzert, 2019c). Für eine erfolgreiche Bearbeitung eines Konzentrationstests müssten diese Teilhandlungen fortlaufend koordiniert, also effizient organisiert und zeitlich abgestimmt werden, was mit einer erhöhten mentalen Anstrengung einhergeht (Moosbrugger & Goldhammer, 2006; Schmidt-Atzert et al., 2008; Westhoff, 1995; Westhoff & Hagemeister, 2005). Dieser typische Präsentationsmodus unterscheidet Konzentrationstests von vielen anderen Aufmerksamkeitstests, in denen die Items üblicherweise einzeln und nacheinander dargeboten werden (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, et al., 2012). Bei diesen Aufmerksamkeitstests ist die Geschwindigkeit der Itemdarbietung computergesteuert und automatisiert, so dass nach der Präsentation eines einzelnen Items eine kleine Pause folgt (Interstimulus-Intervall), bevor das nächste Item erscheint. Nach Westhoff (1995) könnten diese kurzen Pausen genutzt werden, um den Akku, zumindest teilweise, aufzuladen. Sie wären daher nicht geeignet, um die Konzentrationsfähigkeit zu erfassen.



### 3.3 Fazit

Die Gruppe der Konzentrationstests ist recht heterogen und umfasst Durchstreich-, simple arithmetische, Sortier- und Transformationsaufgaben (Büttner & Schmidt-Atzert, 2004). Dennoch teilen diese Tests wesentliche Gemeinsamkeiten, wie relativ simple Aufgabenstellungen und einen charakteristischen Präsentationsmodus, bei dem viele Stimuli gleichzeitig dargeboten werden und kontinuierlich zu bearbeiten sind (Westhoff & Hagemester, 2005). In diesem Präsentationsmodus spiegeln sich einige Schlüsselemente der Konzentration wider, da die kontinuierliche Bearbeitung vieler Stimuli hohe energetische Anforderungen stellt und außerdem erfordert, dass die einzelnen, zur Aufgabenerfüllung nötigen Teilhandlungen fortlaufend koordiniert und kontrolliert werden (Westhoff, 1995).

## **4. Korrelative Validitätsbelege für Konzentrationstests**

Es gibt eine ganze Reihe von Untersuchungen zur psychometrischen Güte verschiedener Konzentrationstests. Die Durchführungs-, Auswertungs- und Interpretationsobjektivität hängen dabei unter anderem vom Detaillierungsgrad und der Qualität des Manuals der einzelnen Konzentrationstests ab. Ähnliches gilt für die Normen. Für einige moderne, viel eingesetzte Konzentrationstests wie den d2-R oder den FAIR-2 können sie sicherlich als gut gelten (Daseking & Putz, 2015; Petermann, 2011). Zahlreiche Untersuchungen berichten auch über sehr zufriedenstellende Reliabilitäten verschiedener Konzentrationstests, sowohl in Bezug auf die interne Konsistenz als auch in Bezug auf Testwiederholungen (Moosbrugger & Oehlschlägel, 2011; Schmidt-Atzert & Brickenkamp, 2017; Steinborn, Langner, Flehmig, & Huestegge, 2017). Das besondere Interesse dieser Arbeit gilt allerdings dem Königs-kriterium der Güte eines Tests, nämlich seiner Validität, mit der sich die folgenden Abschnitte befassen. Dafür werden im Folgenden zunächst korrelative Validitätsbelege zur Konstrukt- und Kriteriumsvalidität dieser Tests zusammengetragen.

### **4.1 Konstruktvalidität – Konvergente Validität**

Wie bereits im Abschnitt zu den Charakteristika von Konzentrationstests angemerkt wurde, ist diese Gruppe von Leistungstests eher heterogen, so unterscheiden sich verschiedene Konzentrationstests oft erheblich in Bezug auf das Stimulusmaterial, das verwendet (z.B. figural, numerisch oder verbal) und die Aufgabe (z.B. Zielreize durchstreichen, Rechnen, Sortieren), die an die Testperson gestellt wird (Büttner & Schmidt-Atzert, 2004). Dennoch konnte in einer Reihe von Untersuchungen für die Tempowerte in verschiedenen Konzentrationstests gezeigt werden, dass diese untereinander hoch korrelieren und in exploratorischen und auch konfirmatorischen Faktorenanalysen einen gemeinsamen Faktor bilden, was für ihre Konstruktvalidität spricht (Mirsky et al., 1991; Schmidt-Atzert et al., 2006; Schmidt, Trueblood, Merwin, & Durham, 1994). Beispielsweise setzten Schmidt-Atzert et al. (2006) eine ganze Reihe von Konzentrationstests sowie Tests der Verarbeitungsgeschwindigkeit ein, die die Kriterien von üblichen Konzentrationstests erfüllen, aber unter anderer Bezeichnung Verwendung finden. Darunter war der d2 (Brickenkamp, 2002), das Frankfurter Aufmerksamkeitsinventar (FAIR; Moosbrugger & Oehlschlägel, 1996), der Revisions-Test (Marschner, 1980), Subtests aus dem Berliner Intelligenzstrukturtest (BIS; Jäger, Süß, & Beauducel, 1997), wie zum Beispiel der Zahlen-Symbol Test und viele weitere Tests. Besonders wichtig ist, dass sich zwar Faktoren für die

unterschiedlichen Inhalte dieser Tests, also für verbales, numerisches und figürliches Testmaterial herausbildeten, diese Tests aber dennoch auf einem gemeinsamen Faktor luden. Die Ergebnisse sprechen dafür, dass diese kognitiven Tests trotz ihrer Heterogenität eine gemeinsame Fähigkeit erfassen, die von den Autoren Konzentration genannt wird. Unter anderem auf der Basis dieser Befunde wird im Rahmen der vorliegenden Arbeit ein gemeinsames Prozessmodell für Konzentrationstests vorgeschlagen (siehe auch *Beitrag 1*, Blotenberg & Schmidt-Atzert, 2019c).

#### 4.2 Konstruktvalidität – Diskriminante Validität

Bereits im Jahre 1904 machte Charles Spearman die fundamentale Beobachtung, dass alle kognitiven Leistungen miteinander korrelieren, ein Phänomen, das als „positive manifold“ in die Geschichte der Intelligenzforschung einging und seither unzählige Male repliziert wurde (siehe auch Horn, 1968 für einen Übersichtsartikel). Das Wissen über dieses Phänomen ist wichtig, um Interkorrelationen verschiedener kognitiver Tests interpretieren zu können, denn selbst bei relativ konstruktfernen Konzepten können üblicherweise keine Nullkorrelationen erwartet werden.

In der Konzentrationsforschung werden diskriminante Validitätsbelege insbesondere zu höheren kognitiven Fähigkeiten erbracht. Die Konzentrationsfähigkeit stellt dabei eine notwendige, aber nicht hinreichende Bedingung für komplexe kognitive Leistungen dar und sollte sich daher von der Leistung in Intelligenz- oder Arbeitsgedächtnistests abgrenzen lassen. Abhängig vom Konzentrationstest und dem Maß der Arbeitsgedächtnisleistung liegen die Korrelationen zwischen den Tempowerten in Konzentrationstests und der Leistung in Arbeitsgedächtnistests im niedrigen bis mittleren Bereich ( $r = .01 - .37$ , Krumm, Lipnevich, Schmidt-Atzert, & Bühner, 2012; Schmidt-Atzert & Ising, 1997).

In einer Metaanalyse, die sechs Studien und insgesamt 6115 Teilnehmer umfasste, fanden sich Korrelationen zwischen  $r = .11$  und  $r = .44$  zwischen den fehlerbereinigten Tempowerten verschiedener Konzentrationstests und der Leistung im Intelligenztest. Der abhängig von der Stichprobengröße gewichtete Mittelwert lag dabei bei .29 (Schmidt-Atzert & Bühner, 2000). Für den fehlerbereinigten Tempowert (Konzentrationsleistungswert, KL) im Konzentrationstest d2-R fand sich über verschiedene Studien hinweg ( $N = 1700$ ) eine mittlere Korrelation von  $r = .32$  mit unterschiedlichen Intelligenztests (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010). In Bezug auf Tests zum schlussfolgernden Denken, die als Kern der

Intelligenz gelten (siehe Carroll, 1993), zeigten sich in einer Untersuchung von Krumm et al. (2009) Zusammenhänge zwischen  $r = .25$  für den Konzentrationstest d2 (Brickenkamp, 2002) und  $r = .33$  für den Revisionstest (Marschner, 1980) mit dem Basismodul des Intelligenzstrukturtests 2000 R (I-S-T 2000 R; Liepmann, Beauducel, Brocke, & Amthauer, 2007).

Einen Sonderfall unter den Intelligenzaufgaben stellen einige Intelligenztests oder Subtests aus Intelligenztestbatterien dar, die die Wahrnehmungs-, Verarbeitungs- oder Bearbeitungsgeschwindigkeit erfassen sollen. Diese Tests, wie beispielsweise der Zahlen-Verbindungs-Test (Oswald & Roth, 1987) oder auch der Zahlen-Symbol Test aus dem Berliner Intelligenzstrukturtest (BIS; Jäger et al., 1997) zeigen oft starke Überschneidungen mit klassischen Konzentrationstests und erfüllen die typischen Kriterien, wie die kontinuierliche Bearbeitung vieler gleichzeitig dargebotener, homogener Stimuli und die Anwendung einer einfachen Regel. Tatsächlich konnten Schmidt-Atzert et al. (2006) zeigen, dass diese Tests auf dem gleichen Faktor laden wie klassische Konzentrationstests und somit die gleiche kognitive Fähigkeit erfassen. Es kommt nun auf die Perspektive an, ob man diese Tests als Tests der Verarbeitungsgeschwindigkeit oder der Konzentrationsfähigkeit auffassen möchte. Aus Sicht der Konzentrationsforschung zeigt sich das konzentrierte Arbeiten in einer effizienten Arbeitsweise, die zu einer schnellen und genauen Bearbeitung dieser Tests führt. Es wird die „Geschwindigkeit des konzentrierten Handelns“ (Westhoff & Hagemeister, 2001, S. 521) erhoben.

#### **4.3 Kriteriumsvalidität**

Schließlich gibt es auch eine ganze Reihe von Untersuchungen, die für die Kriteriumsvalidität von Konzentrationstests sprechen. So wurden für verschiedene Konzentrationstests wie den d2-R, die serielle mentale Additions- und Vergleichsaufgabe (SMACT) oder den Komplexen Konzentrationstest (KKT) Korrelationen zwischen  $r = .11$  und  $r = .42$  mit Schulleistungen gezeigt (Krumm, Lipnevich, et al., 2012; Steinborn, Flehmig, Westhoff, & Langner, 2008; Westhoff & Graubner, 2003). Für die Papierversion des d2-R wurden unter anderem Zusammenhänge mit der Fahreignung und Auswirkungen von Belastungen am Arbeitsplatz bestätigt (Brickenkamp et al., 2010). Außerdem sind die Leistungen im d2 sensitiv für neurotoxische Substanzen und Psychopharmaka (Meyer-Baron, Schaeper, & Seeber, 2002; Peper, 2004). Schließlich zeigten sich auch Zusammenhänge zwischen der Leistung im d2-R und im Frankfurter Aufmerksamkeits-Inventar 2 (FAIR-2) mit

bestimmten Psychopathologien (z. B. Depression, Schizophrenie, ADHS; Brickenkamp et al., 2010; Moosbrugger & Oehlschlägel, 2011).

#### **4.4 Ein Schwachpunkt: Übungseffekte in Konzentrationstests**

Nachdem sich die vorhergehenden Abschnitte mit korrelativen Validitätsbelegen für Konzentrationstests befasst haben, die als sehr zufriedenstellend gelten können, wird im Folgenden ein Schwachpunkt dieser Tests diskutiert: Die relativ großen Übungseffekte. Übungseffekte beschreiben eine Steigerung der Testleistung bei wiederholter Testvorgabe (Lievens, Reeve, & Heggstad, 2007) und sind ein häufiger und gut belegter Befund für viele Tests der kognitiven Leistungsfähigkeit (Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007; Scharfen, Peters, et al., 2018). In einer kürzlich veröffentlichten Metaanalyse lagen die Übungseffekte in kognitiven Leistungstests im Mittel bei .33 Standardabweichungen nach einmaliger Testwiederholung (Scharfen, Peters, et al., 2018). Für Tests der Verarbeitungsgeschwindigkeit, die große Überschneidungen mit Konzentrationstests aufweisen, lagen sie mit .37 Standardabweichungen etwas höher. Weitere Studien zu Übungseffekten in Konzentrationstests, die in der Metaanalyse nicht berücksichtigt wurden, berichten von noch größeren Leistungssteigerungen zwischen einer halben und einer Standardabweichung von der ersten zur zweiten Testung (Bühner et al., 2006; Steinborn et al., 2017; Westhoff, 1989; Westhoff & Dewald, 1990). Dieser Übungseffekt wächst mit wiederholten Testungen weiter an und hatte in einer Studie von Westhoff & Dewald (1990) selbst nach elfmaliger Wiederholung noch kein Plateau erreicht (siehe auch Scharfen, Blum, et al., 2018). Zusätzlich hält der Effekt der Übung lange an. In einer separaten Metaanalyse zu Übungseffekten bei Aufgaben der mentalen Geschwindigkeit („mental speed“), von denen die Gruppe der Verarbeitungsgeschwindigkeitstests sehr nahe an klassischen Konzentrationstests ist (Krumm, Schmidt-Atzert, Michalczyk, & Danthiir, 2008; Schmidt-Atzert et al., 2006), konnte gezeigt werden, dass diese erst nach einem langen Intervall von fünf Jahren vollständig verschwunden sind (Scharfen, Blum, et al., 2018).

Steinborn et al. (2017) argumentierten, dass der Übungseffekt in Konzentrationstests keine dramatische Rolle spiele, weil er die Reliabilität der Tests nicht deutlich vermindern würde. Allerdings sind Übungseffekte dieser Größenordnung für die Individualdiagnostik durchaus nicht unproblematisch. In vielen Anwendungsbereichen von Konzentrationstests, zum Beispiel in der Verkehrspsychologie, für die beispielsweise der d2 ursprünglich entwickelt wurde (Brickenkamp, 2002), aber auch in der Eignungsdiagnostik, haben die Testpersonen naturgemäß kein Interesse daran, gezielte Testvorbereitung oder Übung zuzugeben. Diese

geübten Testpersonen sind gegenüber Ungeübten im Vorteil und haben höhere Chancen, die Fahreignungsprüfung zufriedenstellend zu meistern oder für eine Stelle ausgewählt zu werden. Auch für die Neuropsychologie gilt, dass Maßnahmen, Trainings oder Medikationen nicht allein durch eine prä-post Messung mit einem Konzentrationstest evaluiert werden können. Zusätzlich müsste eine Kontrollgruppe eingesetzt werden, um zu prüfen, ob der Übungseffekt in der Experimentalgruppe, der die Maßnahme, das Training oder die Medikation zuteil wurde, größer ausfällt als in der Kontrollgruppe. Für die Individualdiagnostik gibt es allerdings kaum Möglichkeiten, den Übungseffekt herauszurechnen oder zu berücksichtigen (Bühner et al., 2006; Hagemeister, 2007; Hagemeister, Scholz, & Westhoff, 2002). Zwar gab es Vorschläge, auch Normen für die wiederholte Testdarbietung anzubieten oder die Testpersonen so lange üben zu lassen, bis sie ihr Leistungsmaximum erreicht haben (Bühner et al., 2006), allerdings wären beide Maßnahmen ohne Zweifel sehr aufwändig und hätten eine Reihe negativer Nebeneffekte (z. B. wäre selbst bei Normen für eine Testwiederholung noch immer unklar, wie oft die Testpersonen bereits geübt hatten und wie groß das Retest-Intervall war, die wiederholte Testung könnte zu einem Motivationsverlust der Testperson führen oder zum Verlust des Interesses am Job).

Obwohl die Übungseffekte in kognitiven Leistungstests im Allgemeinen, sowie in Konzentrationstests im Speziellen, sehr gut belegt sind, sind die Ursachen dieses Effekts noch nicht im Detail verstanden (Büttner & Schmidt-Atzert, 2004; Scharfen, Blum, et al., 2018). In der Literatur werden eine ganze Reihe von möglichen Faktoren diskutiert, darunter die Annahme, dass sich die zugrundeliegende Fähigkeit allein durch die Testbearbeitung steigern lässt, aber auch, dass Störfaktoren wie mangelndes Instruktionsverständnis oder Testangst mit der Übung reduziert werden oder dass sich die Testpersonen testspezifische Fähigkeiten und Fertigkeiten angeeignet haben (siehe Lievens, Buyse, & Sackett, 2005; Lievens et al., 2007; Randall & Villado, 2017 für Übersichtsartikel). Auf der Prozessebene ist bislang noch unklar, an welcher Stelle Übungseffekte im Detail auftreten, also welche Prozesse so stark von der Übung profitieren und somit effizienter ablaufen können. An dieser Stelle können experimentelle Studien ansetzen, um besser zu verstehen, welche Prozesskomponenten besonders anfällig für die Effekte der Übung sind. Der dritte Beitrag der vorliegenden Arbeit befasst sich mit dieser Fragestellung. Hier wird der Effekt der Übung auf verschiedene Prozesskomponenten und Mechanismen der Konzentrationstestleistung untersucht (siehe auch *Beitrag 3*, Blotenberg & Schmidt-Atzert, 2019b).

#### 4.5 Fazit

Für etablierte Konzentrationstests wie den d2-R wurden bereits eine ganze Reihe von korrelativen Validitätsbelegen erbracht, die für eine hohe Güte dieser Tests im Hinblick auf ihre Konstrukt- und Kriteriumsvalidität sprechen (z. B. Brickenkamp et al., 2010; Schmidt-Atzert et al., 2006). Ein Schwachpunkt dieser Tests liegt allerdings in den relativ großen Übungseffekten, die in Leistungssituationen wie der Personalauswahl oder der Fahrtauglichkeitsprüfung zu unfairem Wettbewerb führen können und eine prä-post Beurteilung in der Individualdiagnostik unmöglich machen (Hagemeister et al., 2002). Ein tieferes Verständnis der Ursachen dieses Übungseffekts wäre sehr wünschenswert und mithilfe experimenteller Methoden möglich.

## **5. Experimentelle Validitätsbelege für Konzentrationstests**

In den vorherigen Abschnitten wurden wesentliche Definitionselemente der Konzentrationsfähigkeit dargelegt und aufgezeigt, wie diese in Konzentrationstests umgesetzt wurden. Daraufhin wurden umfangreiche korrelative Validitätsbelege angeführt, die für eine hohe Konstrukt- und Kriteriumsvalidität dieser Tests sprechen. Zentrale Annahmen über die Leistung in Konzentrationstests konnten mithilfe dieser Validitätsstudien allerdings nicht geprüft werden: Eine dieser Annahmen wäre, dass der charakteristische Präsentationsmodus von Konzentrationstests, bei dem viele Reize in selbstgewähltem Tempo zu bearbeiten sind, essentiell ist, um den energetischen Aspekt der Konzentrationsfähigkeit abzubilden (Westhoff, 1995). Eine weitere wesentliche Annahme ist, dass beim konzentrierten Arbeiten Teilhandlungen koordiniert und kontrolliert werden müssen und dass es sich bei dieser Koordinationsleistung um eine Komponente der Leistung in Konzentrationstests handelt (Düker, 1957; Westhoff, 1995). Nach Westhoff (1995) sollte ein Konzentrationstest demnach auch verlangen, dass mehrere Teilhandlungen zeitlich organisiert und kontrolliert ausgeführt werden müssen. Um diese Schlüsselannahmen untersuchen zu können, erwies sich die experimentelle Testvalidierung als ein fruchtbarer Ansatz. Eine kleine Auswahl an Studien, die Charakteristika von Konzentrationstests experimentell prüften, soll im Folgenden vorgestellt werden, da sie einen großen Einfluss auf die vorliegende Arbeit hatten.

### **5.1 Experimentelle Studien zur Rolle des Präsentationsmodus**

Zwei experimentelle Untersuchungen befassten sich mit der besonderen Rolle des charakteristischen Präsentationsmodus von Konzentrationstests, in dem viele, gleichzeitig dargebotene Stimuli kontinuierlich und in selbstbestimmtem Tempo zu bearbeiten sind. Wesentliche Merkmale dieses Präsentationsmodus sind zum einen die Stimulusanordnung, also das viele Stimuli gleichzeitig präsentiert werden, und zum anderen das Tempo der Bearbeitung, also das fortlaufend gearbeitet werden muss (Schmidt-Atzert et al., 2008; Schweizer et al., 2005; Westhoff & Hagemeister, 2005). In einer ersten Studie variierten Krumm, Schmidt-Atzert und Eschert (2008) diese Testmerkmale systematisch. So nutzten sie das Aufgabenmaterial verschiedener Aufmerksamkeits- (TAP Subtest Go/NoGo) und Konzentrationstests (Test d2, Revisions-Test) und erstellten für jeden Test computergestützte Versionen, in denen jeweils einer oder viele Reize dargeboten wurden, die entweder in festen Intervallen (mit Interstimulus-Intervall, auch force-paced bzw. computer-paced genannt) dargeboten wurden oder kontinuierlich (ohne Interstimulus-Intervall, auch als self-paced



bezeichnet) zu bearbeiten waren, so dass die Testpersonen entweder kurze Pausen zwischen den Items erhielten oder nicht. Es zeigte sich, dass die Aufmerksamkeits- und Konzentrationstests, die normalerweise moderat miteinander zusammenhingen (z. B.  $r = .34$  oder  $.45$ ), plötzlich hoch miteinander korrelierten, wenn der Präsentationsmodus übereinstimmte (bis  $r = .81$ ). Gleichzeitig hingen die verschiedenen Versionen ein und desselben Tests in den verschiedenen Präsentationsmodi überraschend gering miteinander zusammen. Im Strukturgleichungsmodell kristallisierten sich zum einen Faktoren des Aufgabentyps (d2 und Revisions-Test) und zum anderen Faktoren des Präsentationsmodus (vor allem bezüglich des Tempos; force-paced, self-paced und Papier-Bleistift-Bearbeitung) heraus, die die Testvarianz aufklärten. Diese Ergebnisse sprechen für eine besondere Rolle des Präsentationsmodus und zeigen, dass die Art und Weise der Itemdarbietung die Validität eines Tests stark beeinflussen kann.

In einer weiteren Studie mit ähnlichem experimentellen Aufbau konnte dieses Modell verschiedener Facetten der Aufmerksamkeits- und Konzentrationsleistung bestätigt und die Befunde noch erweitert werden (Krumm, Schmidt-Atzert, et al., 2012). Die Autoren berichteten dabei von einer höheren Korrelation zwischen der Leistung in den Bedingungen mit vielen Stimuli, die in selbstbestimmtem Tempo zu bearbeiten waren und der Leistung in Aufgaben zum schlussfolgernden Denken, als in den übrigen Bedingungen. Sie interpretierten dieses Ergebnis im Sinne einer höheren kognitiven Anforderung dieses Präsentationsmodus, der auch für konventionelle Konzentrationstests charakteristisch ist. Zwar sollte man aus Korrelationen der angegebenen Größenordnung wohl nicht zu starke Schlussfolgerungen ziehen, allerdings gilt die kontinuierliche Bearbeitung vieler gleichzeitig dargebotener Reize auch für viele andere Konzentrationsforscher als vergleichsweise hohe kognitive Anforderung, die mit permanenter mentaler Anstrengung verbunden ist (Schweizer & Moosbrugger, 2004; Schweizer et al., 2005; Westhoff & Hagemeyer, 2005). In diesem Zusammenhang wird auch oft die gesteigerte Koordinationsanforderung genannt, auf die bereits zuvor eingegangen wurde. Dieser haben sich Krumm et al. (2011) in einer weiteren experimentellen Untersuchung gewidmet.

## 5.2 Experimentelle Studien zur Koordination von Teilhandlungen

Krumm et al. (2011) befassten sich auch mit der Rolle der Koordination in Konzentrationstests. Dafür wählten sie den Zahlen-Symbol-Test aus dem Berliner-Intelligenz-Strukturtest (BIS, Jäger et al., 1997), der zwar formal als Test der

Bearbeitungsgeschwindigkeit gilt, de facto aber auch die Kriterien typischer Konzentrationstests erfüllt (Schmidt-Atzert et al., 2006). Bei diesem Test wird zum einen eine Liste mit Zahlen vorgelegt, denen jeweils ein Symbol zugeordnet ist, zum anderen werden Reihen von Zahlen dargestellt, unter denen sich leere Kästchen befinden. Die Aufgabe der Testperson besteht nun darin, die leeren Kästchen mit den passenden Symbolen zu füllen. Die Autoren identifizierten fünf wesentliche Teilhandlungen, die zur Aufgabenerfüllung nötig sind: Die visuelle Suche, das Zeichnen des Symbols, die Transformation der Zahl in ein Symbol, die Inhibition der irrelevanten Zahlen und Symbole und das Merken von Zahl-Symbol-Kombinationen. Sie entwickelten Subtests für diese einzelnen Teilhandlungen und konnten zeigen, dass diese nicht die gesamte Leistung im Zahlen-Symbol-Test vorhersagten, sondern dass ein relativ stabiles Residuum ( $r_{tt} = .44$ , im Retest) zurückblieb, dass sie als Koordinationsfähigkeit interpretierten. Darüber hinaus machten sich die Autoren die vielfach belegten, großen Übungseffekte in Konzentrationstests zunutze. So testeten die Autoren in einer weiteren Studie zwei Gruppen, von denen eine die Teilhandlungen des Zahlen-Symbol-Tests übte und die andere nicht. Anschließend bearbeiteten beide Gruppen die Originalversion des Zahlen-Symbol-Tests. Dabei zeigte sich, dass sich die Leistung zwischen den Gruppen nicht signifikant unterschied und dass der Übungseffekt nach einer erneuten Bearbeitung des Zahlen-Symbol-Tests für beide Gruppen ähnlich groß war. Die Autoren deuteten dieses Ergebnis so, dass der Übungseffekt in Konzentrationstests weniger auf die Übung der einzelnen Teilhandlungen, sondern eher auf weitere Mechanismen, unter anderem auf die Koordination, zurückgehen müsste.

### 5.3 Fazit

Im vorhergehenden Abschnitt wurden empirische Studien vorgestellt, die sich dadurch auszeichnen, dass sie wesentliche theoretische Annahmen zur Konzentrationsfähigkeit und zur Leistung in Konzentrationstests experimentell prüften. Dabei konnte gezeigt werden, dass der Präsentationsmodus einen deutlichen Einfluss auf die Validität von Aufmerksamkeits- und Konzentrationstests hat. Neben der Aufgabe selbst macht es also einen Unterschied, ob in einem Test Einzelstimuli mit kleinen Pausen dargeboten werden, oder ob die Versuchsperson aufgefordert ist, sukzessiv und ohne vorgesehene Pausen auf viele dargebotene Stimuli zu reagieren (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, et al., 2012). Letztere Anforderungen sind typisch für Konzentrationstests und es wurde vielfach argumentiert, dass die selbstgesteuerte Bearbeitung von vielen Stimuli kritisch für die Messung der Konzentrationsfähigkeit ist (Schweizer et al., 2005; Westhoff & Hagemeister,

2005). Somit konnten Kernannahmen der Konzentrationsmessung geprüft und ein wichtiger Beitrag zur experimentellen Validierung dieser Tests geleistet werden. Des Weiteren wurde die bereits von Düker (1957) benannte Koordinationsfähigkeit als wichtige Komponente der Leistung in Konzentrationstests untersucht und erstmalig der Versuch unternommen, diese messbar zu machen. In der Tat konnte nach Auspartialisierung der einzelnen Teilhandlungen aus der Leistung im Gesamttest eine relativ stabile (Koordinations-) Fähigkeit ermittelt werden, die wohl auch zu den Übungseffekten in Konzentrationstests beiträgt (Krumm et al., 2011).

## 6. Zielsetzung des Dissertationsvorhabens

In den ersten Abschnitten dieser Arbeit wurden zentrale Annahmen über die Konzentrationsfähigkeit dargestellt, die Charakteristika von Konzentrationstests beschrieben und eine Reihe korrelativer Validitätsbelege angeführt. Schließlich wurde eine Auswahl experimenteller Studien vorgestellt, denen es gelungen war, einige Schlüsselannahmen über die Leistung in Konzentrationstests zu prüfen. Nichtsdestotrotz bleibt im Feld der Konzentrationsforschung noch vieles unerforscht. So steht ein Prozessmodell bislang noch aus. Für ein besseres Verständnis von Konzentrationstests wäre es allerdings von großer Bedeutung, die Prozesse zu kennen, die bei der Bearbeitung dieser Tests eine wesentliche Rolle spielen. Eine weitere offene Frage betrifft die Übungseffekte in Konzentrationstests. Hier ist noch nicht im Detail verstanden, wie sie verursacht werden und welche Prozesse im Einzelnen von der Übung profitieren (Scharfen, Blum, et al., 2018; Schmidt-Atzert et al., 2004). An dieser Stelle setzt das vorliegende Dissertationsvorhaben an. Ziel ist es, ein generisches Prozessmodell der Leistung in Konzentrationstests vorzuschlagen und empirisch zu prüfen. In einem weiteren Schritt soll auf der Basis dieses Modells untersucht werden, welche der Prozesskomponenten von der Testwiederholung profitieren und effizienter ablaufen. Im Folgenden sollen die Hauptfragestellungen kurz erläutert werden, bevor im nachfolgenden Kapitel die Zusammenfassung der drei einzelnen Manuskripte dieses Dissertationsvorhabens folgt.

### **Auf dem Weg zu einem Prozessmodell:**

#### **Welche Prozesskomponenten leisten einen Beitrag zur Leistung in Konzentrationstests?**

Auf der Basis theoretischer und empirischer Arbeiten soll im ersten Schritt ein Prozessmodell der Leistung in Konzentrationstests vorgeschlagen und untersucht werden. Demnach sollte die Testleistung wesentlich davon abhängen, wie schnell eine Testperson 1) ein Item wahrnimmt, 2) eine simple mentale Operation ausführt, um ein Item zu lösen, 3) motorisch reagiert, um die Lösung zu indizieren und 4) selbstgesteuert zum nächsten Item wechselt (siehe auch Sanders, 1983; Schmidt-Atzert & Brickenkamp, 2017). *Abbildung 3* veranschaulicht diese Annahmen modellhaft.

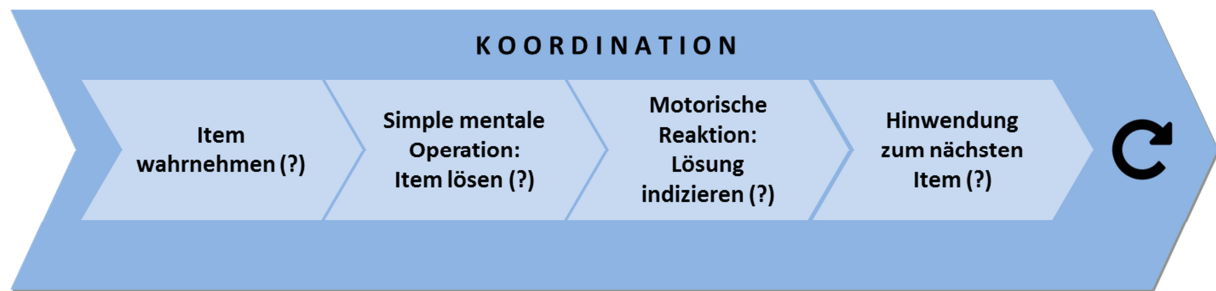


Abbildung 3. Das vorgeschlagene Prozessmodell der Leistung in Konzentrationstests. Demnach hängt die Leistung in Konzentrationstests wesentlich davon ab, wie gut es einer Testperson gelingt, ein Item schnell wahrzunehmen, eine simple mentale Operation auszuführen, um ein Item zu lösen, die richtige Lösung eines Items über eine motorische Reaktion zu indizieren, selbstgesteuert zum nächsten Item zu wechseln und diese einzelnen Prozesskomponenten effizient zu organisieren und zu koordinieren.

Im Rahmen dieser Arbeit wird geprüft, inwieweit sich die vier vorgeschlagenen Prozesskomponenten reliabel messen lassen und ob und in welchem Ausmaß sie die Leistung in Konzentrationstests vorhersagen. Außerdem wird untersucht, inwieweit sich das Prozessmodell – im Sinne der diskriminanten Validität – von Tests höherer kognitiver Fähigkeiten abgrenzen lässt und welche dieser Prozesskomponenten Konzentrationstests mit komplexeren kognitive Leistungstests teilen. Folgende Fragestellungen werden adressiert:

- *Wie stabil sind die postulierten Prozesskomponenten?*
- *Inwieweit können die vier postulierten Prozesskomponenten die Leistung in Konzentrationstests vorhersagen?*
- *Gelingt die Abgrenzung des Prozessmodells von Tests höherer kognitiver Fähigkeiten? Welche Prozesskomponenten teilen diese Tests?*

### **Erweiterung des Prozessmodells:**

#### **Welche weiteren Mechanismen und Fähigkeiten spielen für die erfolgreiche Bearbeitung von Konzentrationstests eine Rolle?**

Im zweiten Schritt soll eine mögliche Erweiterung des Prozessmodells um weitere potentiell relevante Komponenten untersucht werden. Es ist denkbar, dass der charakteristische Präsentationsmodus von Konzentrationstests, bei dem viele Stimuli gleichzeitig präsentiert werden (und fortlaufend zu bearbeiten sind), erhöhte Anforderungen an die Fokussierung des aktuell relevanten Stimulus (unter den vielen Stimuli) stellt. Gleichzeitig könnte der Präsentationsmodus aber auch die Vorschau und eine proaktive Vorverarbeitung nachfolgender Stimuli ermöglichen. Das erweiterte Prozessmodell ist in *Abbildung 4* dargestellt.

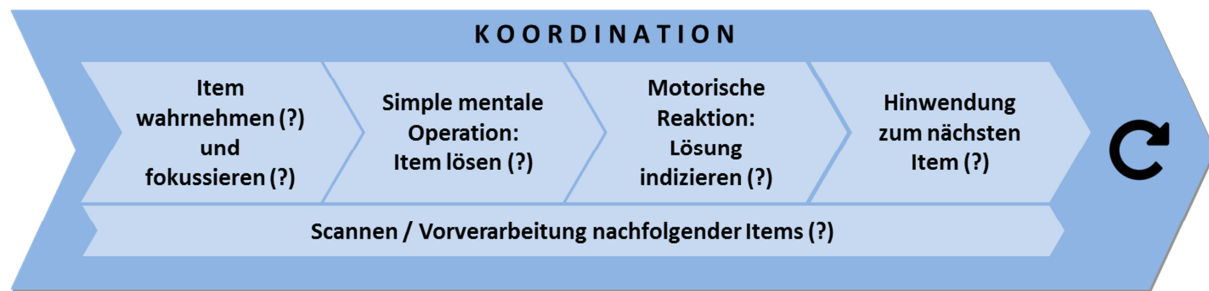


Abbildung 4. Erweiterung des vorgeschlagenen Prozessmodells der Leistung in Konzentrationstests. Demnach hängt die Leistung in Konzentrationstests zusätzlich davon ab, wie gut es der Testperson gelingt, sich auf den aktuell relevanten Stimulus unter zahlreichen Stimuli zu fokussieren sowie die Vorschau auf nachfolgende Stimuli zu nutzen, um sich auf diese vorzubereiten.

Im Rahmen der vorliegenden Arbeit wird untersucht, ob es sich bei der Fokussierung auf den aktuell relevanten Stimulus und bei der proaktiven Vorverarbeitung nachfolgender Stimuli um Komponenten der Leistung in Konzentrationstests handelt. Dafür wird zunächst geprüft, ob sich interindividuelle Unterschiede in diesen angenommenen Komponenten reliabel messen lassen. Schließlich wird geprüft, ob die Fokussierungsleistung und die Vorverarbeitung nachfolgender Stimuli mit der Leistung in Konzentrationstests korrelieren. Folgende Fragestellungen werden behandelt:

*Handelt es sich bei der Fokussierungsleistung um eine stabile Komponente der Leistung in Konzentrationstests?*

- *Inwieweit findet sich ein Nachteil (in Reaktionszeiten und Fehlerraten) für die gleichzeitige Präsentation vieler im Vergleich zu Einzelstimuli?*
- *Gibt es stabile interindividuelle Unterschiede darin, wie gut es den Testpersonen gelingt, sich auf den relevanten Stimulus zu fokussieren?*
- *Hängen interindividuelle Unterschiede in der Fokussierungsleistung mit der Leistung in Konzentrationstests zusammen?*

*Handelt es sich bei der Vorverarbeitung nachfolgender Stimuli um eine stabile Komponente der Leistung in Konzentrationstests?*

- *Inwieweit findet sich ein Vorteil für die Präsentation von Reihen von Stimuli, die im Nachfolgenden relevant werden gegenüber der Darbietung von vielen Stimuli, die keine Vorverarbeitung nachfolgender Stimuli erlauben?*
- *Gibt es stabile interindividuelle Unterschiede darin, wie gut es den Testpersonen gelingt, nachfolgende Stimuli vorzuverarbeiten?*

- *Hängen interindividuelle Unterschiede in der Vorverarbeitung nachfolgender Stimuli mit der Leistung in Konzentrationstests zusammen?*

### **Welche der Prozesskomponenten profitieren in welchem Ausmaß von der Übung?**

Im letzten Schritt sollen die gewonnenen Erkenntnisse über die Prozesskomponenten der Leistung in Konzentrationstests genutzt werden, um dem Übungseffekt in Konzentrationstests näher auf den Grund zu gehen. Ziel ist es, auf der Prozessebene besser zu verstehen, wie die großen Übungseffekte in Konzentrationstests entstehen. Dafür wird untersucht, inwieweit die einzelnen Prozesskomponenten von der Übung profitieren und durch die Testwiederholung schneller ablaufen. Folgende Fragestellungen werden bearbeitet:

*Welche Prozesskomponenten profitieren von der Übung und in welchem Ausmaß?*

- *Inwieweit profitieren die Wahrnehmungsprozesse, die simple mentale Operation zur Lösung des Items und die motorischen Prozesse von der Übung und laufen schneller ab?*
- *Gibt es eine Interaktion der Übung mit dem Tempo der Darbietung in dem Sinne, dass der selbstgesteuerte Wechsel zwischen den Items durch die Übung effizienter wird?*
- *Findet sich eine Interaktion der Übung mit der Stimulusanordnung in dem Sinne, dass es den Testpersonen durch die Übung besser gelingt, die nachfolgenden Stimuli vorzuverarbeiten?*

## 7. Empirischer Teil

### 7.1 Auf dem Weg zu einem Prozessmodell für Konzentrationstests (Beitrag 1)

Blotenberg, I., & Schmidt-Atzert, L. (2019c). Towards a process model of sustained attention tests. *Journal of Intelligence*, 7(1), 3, doi: 10.3390/jintelligence7010003.

Die Originalarbeit zu dieser Zusammenfassung findet sich in Anhang A.

#### *Einführung*

Der erste Beitrag befasst sich mit den Prozesskomponenten bei der Bearbeitung von Konzentrationstests. Dabei werden theoretische Überlegungen sowie frühere Forschungsergebnisse zu diesen Tests berücksichtigt. Wie bereits im theoretischen Hintergrund dieser Arbeit dargelegt, ist die Gruppe der Konzentrationstests recht heterogen (Schmidt-Atzert & Amelang, 2012; Westhoff, 1995). Allerdings gibt es auch einige charakteristische Gemeinsamkeiten, wie die Einfachheit von Stimulusmaterial und Aufgaben und der Präsentationsmodus, der durch die fortlaufende Reaktion auf einfach wahrnehmbare, recht homogene Stimuli gekennzeichnet ist (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, et al., 2012; Westhoff & Hagemeyer, 2005). Tatsächlich zeigen Studien zur Konstruktvalidität, dass verschiedene Konzentrationstests recht hoch miteinander zusammenhängen, einen gemeinsamen Faktor bilden und somit etwas Gemeinsames erfassen (Mirsky et al., 1991; Schmidt-Atzert et al., 2006). Das Ziel der vorliegenden Untersuchung bestand nun darin, ein generisches Prozessmodell für diese Gruppe von Leistungstests vorzuschlagen und zu validieren, da es bisher noch keine prozessorientierten Modelle der Konzentrationsleistung gibt.

Dabei wurden vier wesentliche Prozesskomponenten postuliert, die in konventionellen Konzentrationstests eine Rolle spielen sollten: Zunächst verlangen alle Tests 1) die Wahrnehmung der Items (Wahrnehmungsgeschwindigkeit) und daraufhin 2) eine einfache mentale Operation, um das Item zu lösen (Geschwindigkeit der mentalen Operation). Die geforderte mentale Operation unterscheidet sich dabei zwischen den Tests, sie kann beispielsweise den Abgleich mit einem Zielreiz, eine simple Rechenoperation oder die Zuordnung zu einer Kategorie beinhalten, ist aber in jedem Fall kognitiv niederschwellig. Nun muss 3) eine motorische Reaktion erfolgen, zum Beispiel das Markieren, Durchstreichen oder Setzen eines Häkchens bei Tests im Papier-Bleistift-Format oder auch ein Mausklick bei



elektronischen Konzentrationstests (motorische Geschwindigkeit). Zu guter Letzt muss, bedingt durch die kontinuierliche Bearbeitung dieser Tests, 4) ein selbstgesteuerter Wechsel zum nächsten Item stattfinden (siehe auch Sanders, 1983; Schmidt-Atzert & Brickenkamp, 2017). Die Leistung in Konzentrationstests sollte demnach maßgeblich von der Geschwindigkeit dieser vier Prozesskomponenten abhängen. Gleichzeitig sollten die Prozesskomponenten unzureichend sein, um große Varianzanteile in Aufgaben höherer kognitiver Fähigkeiten, wie Aufgaben zum schlussfolgernden Denken oder der Arbeitsgedächtniskapazität, aufzuklären. Für die erfolgreiche Bewältigung dieser komplexeren Aufgaben sollten zusätzliche, anspruchsvollere mentale Operationen nötig sein, die in unserem Prozessmodell der Leistung in Konzentrationstests nicht abgebildet sind (siehe auch Hunt, 1980; Sternberg, 1977; Unsworth & Engle, 2006, 2007).

### ***Fragestellungen***

- *Wie stabil sind die postulierten Prozesskomponenten?*
- *Inwieweit können die vier postulierten Prozesskomponenten die Leistung in Konzentrationstests vorhersagen?*
- *Gelingt die Abgrenzung des Prozessmodells von Tests höherer kognitiver Fähigkeiten? Welche Prozesskomponenten teilen diese Tests?*

### ***Methode***

Zur Validierung des Prozessmodells sollten Aufgaben ausgewählt werden, die die interessierenden Komponenten abbilden können. Es ist allerdings kaum möglich, Prozesskomponenten gänzlich isoliert zu erfassen (siehe auch Neubauer & Knorr, 1997; Townsend, 1990). Daher wurden geeignete Aufgaben gesucht, in denen die interessierende Prozesskomponente eine besonders große Rolle spielte. Hierbei fiel die Wahl auf die Inspection Time Task (Vickers, Nettelbeck, & Willson, 1972) und die einfache Reaktionszeitaufgabe, die in der Literatur als klassische Maße für die Wahrnehmungsgeschwindigkeit (Brand & Deary, 1982; Schweizer & Koch, 2003) beziehungsweise die motorische Geschwindigkeit (Ackerman, 1988; Fleishman & Hempel, 1954; Sternberg, Monsell, Knoll, & Wright, 1978) gelten. Allerdings stellt die einfache Reaktionszeitaufgabe neben motorischen auch geringfügige Wahrnehmungsanforderungen (Neubauer & Knorr, 1997), weshalb die Inspection Time Task im Strukturgleichungsmodell aus der einfachen Reaktionszeitaufgabe auspartialisiert wurde, um ein Maß für die motorische

Geschwindigkeit zu erhalten. Die Messung der Reaktionen auf einzelne Items ist bei der Darbietung von konventionellen Konzentrationstests leider nicht möglich, da nur ein Gesamtwert der bearbeiteten Items ermittelt werden kann. Daher wurde eine modifizierte Version des gut etablierten Konzentrationstests d2-R erstellt, um die Geschwindigkeit der mentalen Operation zu messen. Da diese Aufgabe zusätzlich Anforderungen an die Wahrnehmungs- sowie die motorische Geschwindigkeit stellt, wurden diese Maße im Strukturgleichungsmodell auspartialisiert (siehe auch Neubauer & Knorr, 1997). Zusätzlich wurde beim modifizierten d2 die Zeit zwischen der Darbietung der Items, das sogenannte Reaktions-Stimulus-Intervall, variiert, um zu untersuchen, ob es stabile interindividuelle Unterschiede in Bezug auf die Zeit gibt, die von einer Testperson für den Itemwechsel benötigt wird oder in Bezug auf Kosten, die entstehen, wenn nicht ausreichend Zeit zwischen der Darbietung der einzelnen Items vorhanden ist. Dabei wurde angenommen, dass die Leistung leidet, wenn die Intervalle zwischen den Items zu kurz sind, während sie sich bei längeren Intervallen verbessern sollte (Jentsch & Dudschig, 2009; Rabbitt, 1969).

Um die Leistungen in diesen Aufgaben zur Leistung in konventionellen Konzentrationstests in Beziehung setzen zu können, wurden ein figuraler (der d2-R elektronisch; Schmidt-Atzert & Brickenkamp, 2017), ein numerischer (der Revisions-Test; Marschner, 1980) und ein verbaler Konzentrationstest (drei Subtests des Berliner Intelligenzstrukturtests zur verbalen Bearbeitungsgeschwindigkeit; Jäger et al., 1997) eingesetzt. Zusätzlich wurden, zur Untersuchung der diskriminanten Validität, auch Tests der höheren kognitiven Leistungsfähigkeit aufgenommen. Zum Einsatz kamen drei Subtests zum schlussfolgernden Denken (Matrizen, Zahlenfolgen und Analogien aus dem I-S-T 2000-R; Liepmann et al., 2007) und drei Tests der Arbeitsgedächtniskapazität (Oswald, McAbee, Redick, & Hambrick, 2014), die ebenfalls jeweils figurales, numerisches und verbales Aufgabenmaterial enthielten.

### ***Ergebnisse***

Es konnte anhand zweier Studien ( $N_{\text{Studie 1}} = 103$ ,  $N_{\text{Studie 2}} = 100$ ) mittels Strukturgleichungsmodellierungen gezeigt werden, dass die Geschwindigkeit in den einzelnen Prozesskomponenten die Leistung in den Konzentrationstests maßgeblich vorhersagte. Vor allem die Wahrnehmungsgeschwindigkeit und die Geschwindigkeit der einfachen mentalen Operation waren starke Prädiktoren, während sich für die motorische Geschwindigkeit ein Trend für einen kleineren Einfluss auf die Testleistung zeigte (Studie 1:  $p = .13$ , Studie 2:  $p = .08$ ). Zwei unterschiedliche Operationalisierungen der Itemwechselzeit resultierten in

unreliablen Maßen, die nicht mit der Leistung in Konzentrationstests oder anderen Leistungstests zusammenhingen. Die Abgrenzung des Prozessmodells von höheren kognitiven Leistungen gelang weitgehend, dabei zeigten sich die Prozesskomponenten vorhersageschwach für die Tests der Arbeitsgedächtniskapazität, waren aber in Studie 1, anders als erwartet, gute Prädiktoren für die Leistung im schlussfolgernden Denken. In der zweiten Studie zeigte sich erwartungskonform, dass die Prozesskomponenten nur einen kleineren Anteil der Varianz in den Aufgaben zum schlussfolgernden Denken vorhersagen konnten. Vor allem die Wahrnehmungsgeschwindigkeit zeigte sich hier als guter Prädiktor der Testleistung.

### ***Diskussion***

Im ersten Beitrag wurde, auf der Basis früherer Untersuchungen, ein generisches Prozessmodell der Leistung in Konzentrationstests vorgeschlagen und validiert. Dabei waren vor allem die Prozesskomponenten der Wahrnehmungsgeschwindigkeit und der simplen mentalen Operation vorhersagestark für die Testleistung, während sich ein Trend für einen kleineren Einfluss der motorischen Geschwindigkeit zeigte. Für die vierte postulierte Prozesskomponente, den selbstgesteuerten Wechsel zum nächsten Item, war eine reliable Messung nicht gelungen und es wurde auch kein Zusammenhang mit der Leistung in Konzentrationstests gefunden. Die Abgrenzung des Prozessmodells für Konzentrationstests von höheren kognitiven Fähigkeiten, wie der Arbeitsgedächtniskapazität und dem schlussfolgernden Denken, gelang weitgehend. Es zeigte sich, dass diese komplexeren Aufgaben einzelne Prozesskomponenten mit Konzentrationstests teilen (beispielsweise Anforderungen an schnelle Wahrnehmungsprozesse in Speed-Tests zum schlussfolgernden Denken), aber auch, dass das Prozessmodell unzureichend war, um den komplexen mentalen Operationen in diesen Aufgaben Rechnung zu tragen.

Insgesamt konnte im ersten Beitrag das vorgeschlagene Prozessmodell für Konzentrationstests zumindest in drei wesentlichen Komponenten bestätigt werden. Es ist somit das, nach unserem Kenntnisstand, erste vorgeschlagene und empirisch untersuchte Prozessmodell für diese wichtige Gruppe der allgemeinen Leistungstests. Eine große Frage blieb allerdings offen: Der Prozesskomponente des selbstgesteuerten Itemwechsels kam in der vorliegenden Untersuchung insofern eine wichtige Rolle zu, als dass sie das fortlaufende, selbstgesteuerte Arbeiten abbilden sollte, das für Konzentrationstests so charakteristisch ist. Für diese Komponente konnten in der vorliegenden Untersuchung aber keine stabilen

interindividuellen Unterschiede gefunden werden. Zukünftige Studien sollten sich der Frage widmen, wie sich die selbstgesteuerte, fortlaufende Bearbeitung vieler Items auf die Informationsverarbeitung in Konzentrationstests auswirkt.

**Anteil des Autors an Manuskript #1:** Die Autorin dieser Dissertation hat beide Studien mit der Unterstützung von Prof. Dr. Schmidt-Atzert geplant, modifizierte Versionen des Konzentrationstests d2-R programmiert, die Datenerhebung organisiert, die Daten mit erhoben, die Daten ausgewertet, das gesamte Manuskript verfasst und die Anmerkungen von Prof. Dr. Schmidt-Atzert eingearbeitet. Das Manuskript wurde am 28.01.2019 im Peer-Review-Journal *Journal of Intelligence* veröffentlicht. Der Eigenanteil der Autorin an diesem Manuskript beträgt 75 %.

## **7.2 Über die Charakteristika der Konzentrationsleistung - Der Vorteil der Vorschau (Beitrag 2)**

Blotenberg, I., & Schmidt-Atzert, L. (2019a). *On the characteristics of sustained attention test performance - The role of the preview benefit*. Manuskript eingereicht.

Die Originalarbeit zu dieser Zusammenfassung findet sich in Anhang B.

### ***Einführung***

Die zweite Untersuchung befasst sich mit dem charakteristischen Bearbeitungsmodus von Konzentrationstests und den spezifischen kognitiven Anforderungen, die damit einhergehen. Es wurde bereits gezeigt, dass der Präsentationsmodus einen Einfluss auf die Validität von Aufmerksamkeits- und Konzentrationstests hat (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, et al., 2012). Der Präsentationsmodus von Konzentrationstests zeichnet sich dabei durch zwei wesentliche Merkmale aus: Erstens werden viele Stimuli gleichzeitig dargeboten, und zweitens sind diese kontinuierlich und in selbstbestimmtem Tempo zu bearbeiten bis die Testzeit abgelaufen ist. Dieser Modus sollte von der Testperson zum einen verlangen, dass sie jeweils den Stimulus unter den vielen Stimuli fokussiert, der zu diesem Zeitpunkt relevant ist. Tatsächlich sind die kognitiven Kontrollanforderungen bei der Darbietung flankierender Stimuli in der Kognitionspsychologie vielfach untersucht worden (Eriksen, 1995; White, Ratcliff, & Starns, 2012). In diesem Beitrag wird daher geprüft, inwieweit sie für die Leistung in Konzentrationstests eine Rolle spielen.

Zudem sollte die Darbietung vieler Stimuli auch die Vorschau und damit die proaktive Vorverarbeitung nachfolgender Stimuli ermöglichen. Aus der Leseforschung ist bekannt, dass sich die Darbietung vieler Stimuli nicht ausschließlich störend, sondern sogar positiv auf die Verarbeitung auswirken kann, wenn diese Stimuli im weiteren Verlauf relevant werden. Belege dafür stammen vor allem aus Eyetracking-Studien, bei denen sich wiederholt zeigte, dass Testpersonen während der Fixation eines Wortes bereits das nachfolgende Wort, zumindest in Ansätzen, verarbeiteten, so dass sie später signifikant schneller darauf reagieren konnten (für Überblicksartikel siehe Rayner, 1998, 2009; Schotter, Angele, & Rayner, 2012). Tatsächlich profitieren auch einige alltägliche Aufgaben, die Konzentration verlangen, wie zum Beispiel das Fahren eines Autos oder die Suche des richtigen Weges in einer Großstadt, von der proaktiven Verarbeitung potentiell relevanter Stimuli in der Umgebung. In diesem

Beitrag soll daher untersucht werden, ob die Fähigkeit zur proaktiven Vorverarbeitung nachfolgender Stimuli in Konzentrationstests zum Tragen kommt.

Drittens sollte die kontinuierliche Bearbeitung der Items auch den selbstgesteuerten Wechsel zwischen den Items erforderlich machen (siehe auch Schmidt-Atzert & Brickenkamp, 2017). Diese Anforderung wurde bereits im Rahmen des Prozessmodells im ersten Beitrag untersucht und wird im Folgenden nur im Rahmen potentieller Interaktionen zwischen der Fokussierung, der Vorverarbeitung nachfolgender Stimuli und dem Itemwechsel betrachtet.

### ***Fragestellungen***

*Handelt es sich bei der Fokussierungsleistung um eine stabile Komponente der Leistung in Konzentrationstests?*

- *Inwieweit findet sich ein Nachteil (in Reaktionszeiten und Fehlerraten) für die gleichzeitige Präsentation vieler im Vergleich zu Einzelstimuli?*
- *Gibt es stabile interindividuelle Unterschiede darin, wie gut es den Testpersonen gelingt, sich auf den relevanten Stimulus zu fokussieren?*
- *Hängen interindividuelle Unterschiede in der Fokussierungsleistung mit der Leistung in Konzentrationstests zusammen?*

*Handelt es sich bei der Vorverarbeitung nachfolgender Stimuli um eine stabile Komponente der Leistung in Konzentrationstests?*

- *Inwieweit findet sich ein Vorteil für die Präsentation von Reihen von Stimuli, die im Nachfolgenden relevant werden gegenüber der Darbietung von vielen Stimuli, die keine Vorverarbeitung nachfolgender Stimuli erlauben?*
- *Gibt es stabile interindividuelle Unterschiede darin, wie gut es den Testpersonen gelingt, nachfolgende Stimuli vorzuverarbeiten?*
- *Hängen interindividuelle Unterschiede in der Vorverarbeitung nachfolgender Stimuli mit der Leistung in Konzentrationstests zusammen?*

### ***Methode***

Für diese Fragestellung wurde ein computergestützter, modifizierter d2 mit insgesamt sechs verschiedenen Bedingungen erstellt, in denen zum einen die Stimulusanordnung und zum anderen das Bearbeitungs tempo variiert wurden. Dieser wurde von insgesamt hundert

Versuchspersonen bearbeitet. In Bezug auf die Stimulusanordnung wurden entweder Einzelstimuli, Blöcke von Stimuli oder Reihen von Stimuli dargeboten. Bei der Präsentation von Blöcken von Stimuli war jeweils nur der Mittlere relevant, während die beiden flankierenden Stimuli als Distraktoren dienten. Dabei wurde ausbalanciert, ob einer, beide oder keiner der Distraktoren mit dem relevanten, mittleren Reiz kongruent waren (also die gleiche Reaktion erforderlich machte). Im Gegensatz zur Bedingung mit Blöcken von Stimuli wurde, bei der Darbietung von Reihen von Stimuli, der jeweils nachfolgende Stimulus sukzessiv relevant. Es wurde angenommen, dass die Darbietung einzelner Stimuli geringe Anforderungen an die Fokussierung stellen sollte, während die Darbietung von Blöcken und Reihen von Stimuli mit einer erhöhten Fokussierungsanforderung einhergehen sollte. Allerdings ermöglichte die Darbietung von Reihen von Stimuli eine Vorschau auf die nachfolgenden Stimuli und eine proaktive Vorverarbeitung, während das bei Blöcken von Stimuli nicht möglich war.

Um auch potentielle Interaktionen mit dem Itemwechsel untersuchen zu können, wurde zusätzlich das Bearbeitungstempo manipuliert. So folgte auf eine Itemantwort entweder ein kurzes Reaktions-Stimulus-Intervall (auch als force-paced Bedingung bezeichnet) oder es erschien unmittelbar das nächste Item (auch als self-paced Bedingung bezeichnet). Dabei sollte die letztere Bedingung höhere Anforderungen an den selbstgesteuerten Itemwechsel stellen, weil hier kontinuierlich gearbeitet werden musste. Die sechs Bedingungen waren demnach: (1) Einzelstimuli, force-paced, (2) Einzelstimuli, self-paced, (3) Blöcke von Stimuli, force-paced, (4) Blöcke von Stimuli, self-paced, (5) Reihen von Stimuli, force-paced, (6) Reihen von Stimuli, self-paced. Zusätzlich wurden drei konventionelle Konzentrationstests eingesetzt (d2-R elektronisch, Schmidt-Atzert & Brickenkamp, 2017; Revisions-Test, Marschner, 1980; drei Subtests des BIS zur verbalen Bearbeitungsgeschwindigkeit, Jäger et al., 1997), um die Fokussierungsleistung und die Vorverarbeitung nachfolgender Stimuli in Bezug zur Leistung in konventionellen Konzentrationstests setzen zu können.

## ***Ergebnisse***

Weder in den Reaktionszeiten noch in den Fehlerraten fanden sich signifikante Unterschiede zwischen der Darbietung von Einzelreizen und Blöcken von Reizen. Diese Ergebnisse sprechen nicht dafür, dass es geglückt war, eine erhöhte Fokussierungsanforderung in der Bedingung mit Blöcken von Stimuli zu induzieren.

Dementsprechend war die Reliabilität eines Indexes, der die individuelle Fokussierungsleistung ausdrücken sollte, nur unzureichend ( $r_{tt} = .24$  im Retest). Dieser Index wurde als das Verhältnis der Reaktionszeiten bei der Darbietung von Einzelstimuli zu den Reaktionszeiten bei der Darbietung von Blöcken von Stimuli berechnet und drückte somit aus, wie gut es der Testperson gelang, die erhöhte Fokussierungsanforderung zu kompensieren.

Gleichzeitig fand sich sowohl in den Reaktionszeiten als auch in den Fehlerraten eine deutliche Leistungsverbesserung für die Darbietung von Reihen von Stimuli gegenüber der Darbietung von Einzel- oder Blöcken von Stimuli. Dieser Vorteil spricht dafür, dass die Versuchspersonen die Vorschau auf die nachfolgenden Stimuli genutzt und diese vorverarbeitet hatten. Außerdem fand sich eine Interaktion der Stimulusanordnung mit dem Bearbeitungstempo: Der Vorteil der Darbietung von Reihen von Stimuli war besonders groß, wenn das Bearbeitungstempo kurze Intervalle zwischen aufeinanderfolgenden Stimuli erlaubte (also force-paced war). Dieser Effekt deutet daraufhin, dass die kurzen Intervalle zwischen aufeinanderfolgenden Stimuli (in denen der Bildschirm weiß geworden war) genutzt werden konnten, um die Reaktion auf den nachfolgenden Stimulus weiter vorzubereiten (möglicherweise durch ein visuelles Nachbild, siehe auch Vickers & Smith, 1986). Ein Index, der den individuellen Vorteil durch die Vorverarbeitung nachfolgender Stimuli angab, zeigte sich als sehr reliabel, auch in Bezug auf die Testwiederholung ( $r_{tt} = .81$ ). Er wurde berechnet als das Verhältnis der Reaktionszeiten für Blöcke von Stimuli im Vergleich zu Reihen von Stimuli und drückte somit aus, wie sehr die Leistung einer Testperson von der Möglichkeit des Vorausarbeitens profitierte. Wichtig war zudem, dass interindividuelle Unterschiede in der Vorverarbeitung nachfolgender Stimuli systematisch mit der Leistung in drei Konzentrationstests zusammenhingen, darunter nicht nur mit dem d2-R (elektronisch;  $r = .51$ ), sondern auch mit dem numerischen Revisions-Test ( $r = .35$ ) und verbalen Subtests aus dem BIS ( $r = .26$ ), wobei die letzten beiden im Papier-Bleistift-Format vorgegeben worden waren.

## ***Diskussion***

Im zweiten Beitrag sollte untersucht werden, inwieweit die Fokussierung auf den aktuell relevanten Stimulus und eine proaktive Vorverarbeitung nachfolgender Stimuli eine Rolle für die Leistung in Konzentrationstests spielen. In Bezug auf die Fokussierungsleistung konnte dabei kein Effekt gefunden werden. So unterschieden sich die Bedingungen, die höhere



Fokussierungsanforderungen stellen sollten, nicht signifikant von den Bedingungen, in denen Einzelstimuli dargeboten wurden. Ein Maß für die individuelle Fokussierungsleistung war unreliabel und hing nicht mit der Leistung in Konzentrationstests zusammen. Diese Ergebnisse sind wahrscheinlich darauf zurückzuführen, dass die gestellten Fokussierungsanforderungen in den betreffenden Bedingungen zu niedrig waren. So wurden zwar gleichzeitig drei Stimuli präsentiert, allerdings war dabei immer nur der Mittlere relevant und dieser wurde zusätzlich mit einem Pfeil markiert. Zukünftige Studien sollten demnach eine herausforderndere Bedingung für die Untersuchung der Fokussierungsleistung verwenden.

Das wichtigste Ergebnis dieser Studie war allerdings, dass die Darbietung von Reihen von Stimuli zu einem großen Leistungsvorteil gegenüber den anderen Bedingungen führte. Das bedeutet, dass die Testpersonen die Vorschau auf die nachfolgenden Stimuli genutzt und diese vorverarbeitet hatten, was eine deutliche Leistungssteigerung zur Folge hatte. Hier zeigte sich außerdem, dass die Testpersonen die nachfolgenden Stimuli besonders effizient vorverarbeiten konnten, wenn der Bearbeitungsmodus kurze Intervalle zwischen aufeinanderfolgenden Stimuli erlaubte. Diese Interaktion von Stimulusanordnung und Bearbeitungsmodus unterstrich die leistungssteigernde Rolle der Vorverarbeitung nachfolgender Stimuli. Darüber hinaus waren interindividuelle Unterschiede darin, wie gut es den Testpersonen gelang die nachfolgenden Stimuli vorzuverarbeiten, sehr reliabel und zeigten mittlere bis hohe Zusammenhänge mit der Leistung in verschiedenen Konzentrationstests.

Zusammengenommen konnte im zweiten Beitrag mithilfe experimenteller Methoden eine neue Komponente der Leistung in Konzentrationstests gezeigt werden: Die proaktive Vorverarbeitung nachfolgender Stimuli. Ob es sich dabei um eine Strategie oder um eine stabile Fähigkeit handelt, kann noch nicht abschließend beantwortet werden. Die hohe Reliabilität (insbesondere im Retest) spricht allerdings eher für eine stabile Fähigkeit. Handelte es sich um eine Strategie, so sollte diese auch im Verlaufe der Testbearbeitung erlernbar sein, was die Reliabilität verringern würde. Des Weiteren ist anzumerken, dass sich in der vorliegenden Untersuchung ein Zusammenhang zwischen dem Ausmaß der Vorverarbeitung nachfolgender Stimuli mit der Leistung in sehr verschiedenen Konzentrationstests zeigte, die unterschiedliche Aufgabenmaterialien (figural, numerisch, verbal) und mentale Operationen (z.B. Zielreize wegstreichen, im niedrigen Bereich addieren, Buchstaben ergänzen) verlangten und unterschiedliche Formate (elektronisch oder

Papierformat) verwendeten. Daher ist es sehr wahrscheinlich, dass die Komponente der Vorverarbeitung auch in anderen Konzentrationstests oder anderen Tests der Wahrnehmungs-, Bearbeitungs-, oder Verarbeitungsgeschwindigkeit (sofern diese die kontinuierliche Bearbeitung vieler Stimuli verlangen; Krumm, Schmidt-Atzert, Michalczyk, et al., 2008) eine Rolle spielen könnte. Zukünftige Studien sollten sich der Frage widmen, ob es sich bei dieser Komponente der Testleistung um einen inherenten Bestandteil der Konzentrationsfähigkeit oder möglicherweise um eine nicht intendierte Begleiterscheinung der Testgestaltung handelt.

**Anteil des Autors an Manuskript #2:** Die Autorin dieser Dissertation hat diese Studie mit der Unterstützung von Prof. Dr. Schmidt-Atzert geplant, modifizierte Versionen des Konzentrationstests d2-R programmiert, die Datenerhebung organisiert, die Daten mit erhoben, die Daten ausgewertet, das gesamte Manuskript verfasst und die Anmerkungen von Prof. Dr. Schmidt-Atzert eingearbeitet. Das Manuskript wurde am 11.10.2018 zur Veröffentlichung beim Peer-Review-Journal *European Journal of Psychological Assessment* eingereicht. Der Eigenanteil der Autorin an diesem Manuskript beträgt 75 %.

### 7.3 Zu den Ursachen des Übungseffekts in Konzentrationstests (Beitrag 3)

Blotenberg, I., & Schmidt-Atzert, L. (2019b). *On the locus of the practice effect in sustained attention tests*. Manuskript eingereicht.

Die Originalarbeit zu dieser Zusammenfassung findet sich in Anhang C.

#### **Einführung**

Die dritte Untersuchung behandelt ein weiteres, gut dokumentiertes Phänomen beim Einsatz von Konzentrationstests – den relativ großen Übungseffekt bei wiederholter Testung. In einer kürzlich erschienenen Metaanalyse wurde dieser nach einmaliger Testwiederholung mit einer Effektstärke von 0.37 Standardabweichungen beziffert (Scharfen, Peters, et al., 2018), andere Studien konnten beachtliche Übungseffekte zwischen einer halben und bis zu einer Standardabweichung zeigen (Bühner et al., 2006; Westhoff, 1989; Westhoff & Dewald, 1990). Für den Konzentrationstest d2-R, der Gegenstand dieser Untersuchung ist, lag er zwischen zwei Dritteln in früheren Studien (Bühner et al., 2006; Steinborn et al., 2017) und vier-Fünfteln für die neueste, elektronische Version (Schmidt-Atzert & Brickenkamp, 2017) nach einmaliger Testwiederholung. Bislang ist noch unklar, an welcher Stelle die Übungseffekte im Detail auftreten (Büttner & Schmidt-Atzert, 2004; Scharfen, Blum, et al., 2018), also welche Prozessschritte bei der Bearbeitung dieser Tests im Einzelnen von der Übung profitieren. In der vorliegenden Untersuchung legen wir das bereits vorgestellte Prozessmodell der Konzentrationsleistung zugrunde und widmen uns der Frage, ob die Wahrnehmung der Items, die einfache mentale Operation zur Lösung des Items, die motorische Reaktion oder womöglich auch das Wechseln zwischen den Items oder die Vorverarbeitung nachfolgender Items von der Übung profitiert (siehe Beiträge 1 und 2; Blotenberg & Schmidt-Atzert, 2019a, 2019c). Das übergeordnete Ziel der vorliegenden Untersuchung lag darin, die Übungseffekte in Konzentrationstests besser zu verstehen.

#### **Fragestellungen**

*Welche Prozesskomponenten profitieren von der Übung und in welchem Ausmaß?*

- *Inwieweit profitieren die Wahrnehmungsprozesse, die simple mentale Operation zur Lösung des Items und die motorischen Prozesse von der Übung und laufen schneller ab?*
- *Gibt es eine Interaktion der Übung mit dem Tempo der Darbietung in dem Sinne, dass der selbstgesteuerte Wechsel zwischen den Items durch die Übung effizienter wird?*

- *Findet sich eine Interaktion der Übung mit der Stimulusanordnung in dem Sinne, dass es den Testpersonen durch die Übung besser gelingt, die nachfolgenden Stimuli vorzuverarbeiten?*

### **Methode**

Um diesen Fragen auf den Grund zu gehen, wurden eine Reihe kognitiver Aufgaben eingesetzt, die einhundert Versuchspersonen zweimal bearbeiteten. Zum einen wurde ein modifizierter d2 verwendet, dessen Testcharakteristika dem Original-Test sehr nahe kamen, da hier Reihen von Stimuli dargeboten wurden und fortlaufend zu bearbeiten waren (Bedingung mit Reihen von Stimuli, self-paced). Dieser diente als Modell für einen konventionellen Konzentrationstest und wird dementsprechend im Folgenden als Modell-d2 bezeichnet. Zum anderen wurden Aufgaben eingesetzt, die die Geschwindigkeit in den einzelnen Prozesskomponenten von Konzentrationstests erfassen sollten. Hierbei ist zu bedenken, dass kognitive Prozesse nicht unbedingt strikt seriell ablaufen und diese Aufgaben in der Regel mehr als nur einen Prozessschritt erfordern (Neubauer & Knorr, 1997; Townsend, 1990). Das Ziel lag somit darin, Aufgaben zu finden, die zunehmend mehrere der relevanten kognitiven Prozesse verlangen. Hierfür wurde die Inspection Time Task (Vickers et al., 1972) zur Messung der Wahrnehmungsgeschwindigkeit eingesetzt, da sie als prototypische Wahrnehmungsaufgabe gilt (Brand & Deary, 1982; Schweizer & Koch, 2003). Zur Messung der motorischen Geschwindigkeit wurde die simple Reaktionszeitaufgabe eingesetzt, da sie als klassisches Maß der motorischen Geschwindigkeit gilt (Ackerman, 1988; Fleishman & Hempel, 1954; Sternberg et al., 1978). Allerdings erfasst diese Aufgabe neben motorischen auch basale perzeptuelle Prozesse (Neubauer & Knorr, 1997). Zur Messung der Geschwindigkeit einer einfachen mentalen Operation wurde eine vereinfachte Version des modifizierten d2 erstellt, in der Einzelstimuli nacheinander dargeboten wurden. Dabei gab es jeweils halbsekündige Intervalle (Reaktions-Stimulus Intervalle) zwischen aufeinanderfolgenden Stimuli (Bedingung mit Einzelstimuli, force-paced). Diese Aufgabe umfasste somit neben Wahrnehmungs- und motorischen Prozessen zusätzlich die simple mentale Operation zur Lösung des Items, aber keinen der zusätzlichen Prozesse, die bei Konzentrationstests eine Rolle spielen, wie der selbstgesteuerte Wechsel zwischen den Items oder die proaktive Vorverarbeitung nachfolgender Items (siehe auch Blotenberg & Schmidt-Atzert, 2019a, 2019c).

Schließlich wurden noch weitere Varianten des modifizierten d2 erstellt und deren Stimulusanordnung und Bearbeitungstempo systematisch manipuliert, um zu untersuchen, inwieweit der Itemwechsel oder die proaktive Vorverarbeitung von der Übung profitierte. Das Tempo des modifizierten d2 wurde dabei so variiert, dass es entweder keine Intervalle zwischen aufeinanderfolgenden Items gab, so dass die Testteilnehmer sofort zum nächsten Item wechseln mussten (Itemwechselbedingung, self-paced) oder so, dass es halbsekündige Intervalle zwischen aufeinanderfolgenden Items gab, so dass die Teilnehmer kurze Pausen zwischen aufeinanderfolgenden Items erhielten (force-paced-Bedingung). Die Stimulusanordnung wurde zwischen Einzelstimuli, Blöcken von Stimuli und Reihen von Stimuli variiert und nur die Bedingung mit Reihen von Stimuli erlaubte die Vorschau auf nachfolgende Stimuli und deren Vorverarbeitung.

### ***Ergebnisse***

Wie erwartet zeigte sich für den Modell-d2 ein Übungseffekt in der Größenordnung konventioneller Konzentrationstests, nämlich von etwa zwei Dritteln Standardabweichung ( $d = .62$ ). Für die Aufgaben der Wahrnehmungsgeschwindigkeit und der motorischen Geschwindigkeit zeigten sich kleinere bis mittlere Übungseffekte ( $d = .23$  für die Inspection Time Task,  $d = .46$  für die simple Reaktionszeitaufgabe). Der weitaus größte Übungseffekt fand sich in der vereinfachten Version des modifizierten d2 ( $d = 1.02$ ), der neben Wahrnehmungs- und motorischen Prozessen zusätzlich die einfache mentale Operation zur Lösung des Items, also die Entscheidung, ob es sich um ein d2 handelte oder nicht, verlangte. Der Übungseffekt in diesem vereinfachten modifizierten d2 hing auch substanziell mit dem Übungseffekt im Modell-d2 zusammen ( $r = .33$ , nach doppelter Minderungskorrektur sogar  $r = .50$ ). Der selbstgesteuerte Itemwechsel oder die Vorverarbeitung nachfolgender Stimuli profitierten nicht von der Übung. Die Ergebnisse sprechen somit dafür, dass sich die Übung in kleinerem Ausmaß auf die Wahrnehmungs- und motorische Geschwindigkeit, aber vor allem auf die simple mentale Operation auswirkte, also auf die Prozesse, die der Entscheidung über das Item zugrundeliegen.

### ***Diskussion***

Das Ziel der vorliegenden Untersuchung bestand darin, die Übungseffekte in Konzentrationstests besser zu verstehen. Dafür wurde ein Prozessmodell von Konzentrationstests zugrundegelegt und untersucht, welche Prozesskomponenten und

weiteren Mechanismen von der Übung profitierten. Es zeigten sich recht komplexe Auswirkungen der Übung, die zu kleineren Übungseffekten in der Wahrnehmungs- und der motorischen Geschwindigkeit führte und zum größten Übungseffekt in der Aufgabe, die zusätzlich die Lösung des Items verlangte. Der Übungseffekt in dieser Aufgabe hing auch mit dem Übungseffekt im sogenannten Modell-d2 zusammen, einem eigens programmierten Test, bei dem Reihen von Stimuli kontinuierlich bearbeitet werden mussten und der somit als Modell für einen konventionellen Konzentrationstest diente. Weder der selbstgesteuerte Itemwechsel noch die Vorverarbeitung nachfolgender Stimuli wurden durch die Übung effizienter. Zusammengenommen sprechen die Ergebnisse dafür, dass der deutlichste Übungseffekt beim Lösungsprozess des Items auftritt, also bei den vergleichsweise simplen Entscheidungsprozessen, die bei der Bearbeitung von Konzentrationstests nötig sind (z.B. die Entscheidung, ob es sich bei einem Stimulus um einen Zielreiz handelt oder nicht, oder ob zwei Zahlen in der Summe eine dritte Zahl ergeben oder nicht).

Zusätzlich könnte auch die Koordination der Wahrnehmungs- Lösungs- und motorischen Prozesse von der Übung profitiert haben, denn diese wird im vereinfachten modifizierten d2 ebenfalls verlangt. Hinweise auf Übungseffekte in der Koordination der Teilhandlungen, die für die Bearbeitung einer Aufgabe nötig sind, gab es bereits in einer früheren Studie (Krumm et al., 2011). Leider war eine vollständige Trennung der einzelnen Prozesskomponenten in der vorliegenden Untersuchung nicht möglich. Nichtsdestotrotz konnten erste wichtige Hinweise darüber gesammelt werden, wie der Übungseffekt in Konzentrationstests zustandekommt und es konnte erneut gezeigt werden, dass die psychologische Diagnostik sehr vom Einsatz experimenteller Methoden profitieren kann. Ein tieferes Verständnis der Phänomene, die bei diesen Tests beobachtet werden, ist der Schlüssel, um diese zukünftig zu adressieren.

**Anteil des Autors an Manuskript #3:** Die Autorin dieser Dissertation hat die vorliegende Studie mit der Unterstützung von Prof. Dr. Schmidt-Atzert geplant, modifizierte Versionen des Konzentrationstests d2-R programmiert, die Datenerhebung organisiert, die Daten mit erhoben, die Daten ausgewertet, das gesamte Manuskript verfasst und die Anmerkungen von Prof. Dr. Schmidt-Atzert eingearbeitet. Das Manuskript wurde am 11.02.2019 zur Veröffentlichung beim Peer-Review-Journal *Journal of Intelligence* eingereicht. Der Eigenanteil der Autorin an diesem Manuskript beträgt 75 %.

## 8. Gesamtdiskussion

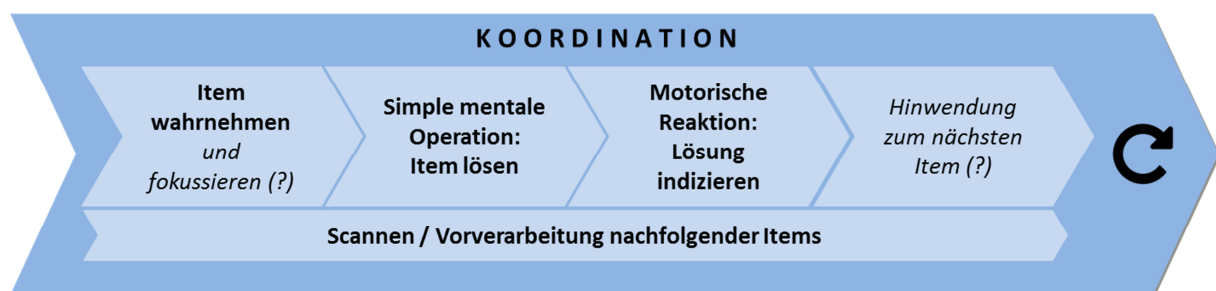
Kognitive Tests zur Messung der Konzentrationsfähigkeit sind seit über einhundert Jahren im Einsatz und erfreuen sich in vielen Anwendungsbereichen der Psychologie großer Beliebtheit (Schmidt-Atzert & Amelang, 2012). Das Ziel der vorliegenden Arbeit lag darin, die Prozesskomponenten zu untersuchen, die bei der Bearbeitung dieser Tests eine wesentliche Rolle spielen, um besser zu verstehen, wie die Leistung in Konzentrationstests zustande kommt. Erkenntnisse über diese Prozesskomponenten sollten außerdem dazu genutzt werden, sich Phänomenen in diesen Tests, wie den relativ großen Übungseffekten, zu nähern.

Dafür wurde im *ersten Beitrag* dieser Arbeit ein generisches Prozessmodell der Leistung in Konzentrationstests vorgeschlagen. Es zeigte sich, dass insbesondere die ersten beiden Prozesskomponenten einen großen Einfluss auf die Leistung in Konzentrationstests hatten, nämlich zum einen 1) die Geschwindigkeit, mit der Items wahrgenommen werden und zum anderen 2) die Geschwindigkeit, mit der eine simple mentale Operation zur Lösung des Items ausgeführt wird (z.B. Abgleich mit Zielreiz, Addition einstelliger Zahlen und Abgleich mit Musterlösung). Außerdem gab es einen Trend für einen kleineren Einfluss der 3) motorischen Geschwindigkeit auf die Leistung in Konzentrationstests. Insgesamt klärten die Prozesskomponenten 55 – 74 % der Varianz der Leistung in den eingesetzten Konzentrationstests auf. Obwohl die Anforderung zum selbstgesteuerten Wechsel zwischen den Items, die für Konzentrationstests üblich ist, die übrigen Prozesskomponenten zu beeinflussen schien, zeigte sich für die vierte angenommene Prozesskomponente, 4) den schnellen selbstgesteuerten Wechsel zwischen den Items, nur eine geringe Stabilität und kein Zusammenhang mit der Testleistung. Im *zweiten Beitrag* wurde das Prozessmodell um eine weitere Komponente erweitert. Die Leistung in Konzentrationstests hängt demnach auch davon ab, wie gut es einer Testperson gelingt, die Vorschau auf nachfolgende Items zu nutzen und diese vorzuverarbeiten. Im *dritten und letzten Beitrag* wurde untersucht, welche der betrachteten Komponenten von der Übung profitieren. Hier zeigten sich kleinere Übungseffekte in der Wahrnehmungs- sowie der motorischen Geschwindigkeit und der weitaus größte Übungseffekt, wenn zusätzlich die Lösung des Items, also eine simple mentale Operation und Entscheidung (z.B. Abgleich mit Zielreiz, Entscheidung ob Zielreiz oder nicht) gefordert war. Weder der Wechsel zwischen den Items noch die proaktive Vorverarbeitung nachfolgender Items wurden durch die Übung effizienter. Im Folgenden werden zunächst die gewonnenen Erkenntnisse über die Prozesskomponenten und anschließend die Übungseffekte

in Konzentrationstests diskutiert. Abschließend soll ein Fazit gezogen und ein Ausblick auf zukünftige Forschung gegeben werden.

### 8.1 Komponenten der Leistung in Konzentrationstests

Bereits Düker und Lienert (1959, S. 3) sahen in der Konzentrationsfähigkeit die „Fähigkeit zur Anspannung zum Zwecke der Koordination“. Dabei definierten sie die Koordinationsleistung als „das zu einer Gesamttätigkeit geordnete Zusammenwirken der Einzeltätigkeiten, die zur Erreichung eines bestimmten Zweckes erforderlich sind“. Jahrzehnte später gelang es Krumm et al. (2011) dieses Schlüsselement der Konzentrationsfähigkeit in einem Konzentrationstest zu zeigen. Das Ziel der vorliegenden Arbeit war es nun zu ergründen, welche Teilhandlungen im Einzelnen bei der Bearbeitung dieser Tests zu koordinieren sind. Auf der Basis theoretischer und empirischer Arbeiten wurde im *ersten Beitrag* ein generisches Prozessmodell der Leistung in Konzentrationstests postuliert und im *zweiten Beitrag* erweitert. Demnach hängt die Leistung einer Testperson in einem typischen Konzentrationstest wesentlich davon ab, wie gut es ihr gelingt 1) ein Item schnell wahrzunehmen und 2) das aktuell relevante Item zu fokussieren (und die Übrigen auszublenden), 3) eine simple mentale Operation schnell auszuführen, um das Item zu lösen, 4) schnell motorisch zu reagieren, um die Lösung zu indizieren, 5) selbstgesteuert zum nächsten Item zu wechseln und 6) parallel zu diesen Prozessen die Vorschau auf nachfolgende Items zu nutzen und diese vorzuverarbeiten. Abbildung 5 stellt die postulierten Prozesskomponenten der Leistung in Konzentrationstests grafisch dar. Im Folgenden sollen die einzelnen Komponenten schrittweise behandelt und ihre Rolle für die Leistung in Konzentrationstests anhand der empirischen Befunde diskutiert werden.



*Abbildung 5.* Die vorgeschlagenen Prozesskomponenten der Leistung in Konzentrationstests. Demnach hängt die Leistung wesentlich davon ab, wie gut es einer Testperson gelingt, ein Item schnell wahrzunehmen, das aktuell relevante Item (unter den vielen präsentierten Items) zu fokussieren, eine simple mentale Operation schnell auszuführen, um das Item zu lösen, die richtige Lösung des Items über eine motorische Reaktion zu indizieren, selbstgesteuert zum nächsten Item zu wechseln, die Vorschau der nachfolgenden Items zu nutzen und diese Items



vorzuverarbeiten und die einzelnen Prozesskomponenten effizient zu koordinieren. Die Komponenten des Prozessmodells, für die sich in den Beiträgen dieser Arbeit empirische Belege finden ließen, sind gefettet, die übrigen Komponenten sind kursiv gedruckt und mit einem Fragezeichen versehen.

### ***8.1.1 Die Rolle der Wahrnehmung***

Die wichtige Rolle der Wahrnehmungsgeschwindigkeit für die Leistung in Konzentrationstests kann als gut gesichert gelten. In den zwei Studien des ersten Beitrags (Blotenberg & Schmidt-Atzert, 2019c) war konsistent gezeigt worden, dass die Wahrnehmungsgeschwindigkeit signifikant und bedeutsam zur Leistung in Konzentrationstests beiträgt. Diese Ergebnisse passen zu einer früheren Studie, die demonstriert hatte, dass der Zusammenhang zwischen Konzentrationstests und Aufgaben des schlussfolgernden Denkens unter anderem auf geteilte Wahrnehmungsanforderungen zurückgeht (Schweizer & Moosbrugger, 2004). Führt man sich den charakteristischen Präsentationsmodus von Konzentrationstests vor Augen, in dem viele Stimuli gleichzeitig präsentiert werden, ist es sicherlich nicht verwunderlich, dass die Leistung in diesen Tests stark von schnellen Wahrnehmungsprozessen profitiert. Einige Tests, die die Kriterien klassischer Konzentrationstests erfüllen, wie beispielsweise die Durchstreichtests „a’s finden“ oder „t’s finden“ werden auch als Tests der „Wahrnehmungsgeschwindigkeit“ eingesetzt (Ackerman, Beier, & Boyle, 2002; Schaie, 1989) und finden sich im einflussreichen CHC-Modell der Intelligenz unter der Komponente „Perceptual Speed (p)“ wieder (McGrew, 2005, 2009). Damit wird explizit auf die bedeutende Rolle schneller Wahrnehmungsprozesse in diesen Tests Bezug genommen.

### ***8.1.2 Die Rolle der Fokussierung***

Die Bedeutung der Fokussierungsleistung in Konzentrationstests bleibt zunächst noch ungeklärt. In der kognitionspsychologischen Forschung sind erhöhte kognitive Kontrollanforderungen durch die Darbietung flankierender Stimuli gut belegt (Eriksen, 1995; Mullane, Corkum, Klein, & McLaughlin, 2009, für Übersichtsartikel). Auch in Konzentrationstests werden viele Stimuli gleichzeitig dargeboten, die erhöhte Kontrollanforderungen stellen sollten. Daher wurde angenommen, dass die Leistung in diesen Tests auch davon abhängen sollte, wie gut es der Testperson gelingt, den aktuell relevanten Stimulus unter den vielen Stimuli zu fokussieren. Der zweite Beitrag der vorliegenden Arbeit befasste sich mit dieser Fragestellung. Allerdings war das ermittelte Maß für die Fokussierungsleistung unreliabel und es fand sich kein Zusammenhang zwischen der

Fokussierungsleistung und der Leistung in konventionellen Konzentrationstests. Es ist wahrscheinlich, dass die verwendeten ablenkenden Stimuli zu schwach waren und die experimentelle Manipulation einer erhöhten Fokussierungsanforderung nicht geglückt war. Deshalb wäre es sinnvoll, die Fokussierungsleistung in Konzentrationstests in einer weiteren Studie mit stärker ablenkenden Reizen erneut zu untersuchen.

### ***8.1.3 Die Rolle der simplen mentalen Operation***

Es kann als gut gesichert gelten, dass die Leistung in Konzentrationstests maßgeblich davon abhängt, wie gut es einer Testperson gelingt, die geforderten simplen mentalen Operationen zur Lösung des Items (z.B. Abgleich mit Zielreiz, Aufaddieren von Zahlen etc.) effizient auszuführen. In den zwei Studien des ersten Beitrags zeigte sich diese Prozesskomponente konsistent als sehr vorhersagestark. In den ersten Kapiteln dieser Arbeit wurde jedoch bereits dargestellt, dass sich Konzentrationstests erheblich in den geforderten mentalen Operation unterscheiden. Ihre Gemeinsamkeit ist, dass die Lösungsprozesse in jedem dieser Tests sehr einfach sind, weshalb diese Tests in verschiedenen Untersuchungen einen gemeinsamen Faktor bildeten (Mirsky et al., 1991; Schmidt-Atzert et al., 2006) und in dieser Arbeit für ein generisches Prozessmodell zusammengefasst wurden. Das heißt, jemand, der schnell Zielreize abgleichen kann, ist in der Regel auch in der Lage, effizient einstellige Zahlen aufzuaddieren oder Wörter zu vervollständigen, bei denen ein Buchstabe fehlt – die verschiedenen Konzentrationstests bilden einen gemeinsamen Faktor. Interessieren allerdings die spezifischen Prozesse bei einem einzelnen Konzentrationstest, so könnten die Lösungsprozesse im Prozessmodell noch weiter ausdifferenziert werden (z.B. Aufaddieren der beiden einstelligen Zahlen, Merken des Ergebnisses, Abgleich mit vorgegebenem Ergebnis beim Revisions-Test). Als Vorbild hierfür könnte eine frühere Untersuchung von Krumm et al. (2011) dienen, in der bereits der Zahlen-Symbol-Test in seine Teilhandlungen zerlegt wurde (siehe auch Gilmore, Royer, Gruhn, & Esson, 2004; Royer, 1971).

### ***8.1.4 Die Rolle der Motorik***

Insgesamt sprechen die Ergebnisse früherer Studien und die aktuellen Befunde dafür, dass die motorische Geschwindigkeit – wenn auch in kleinerem Maße – zur Leistung in Konzentrationstests beiträgt. In den zwei Studien des ersten Beitrags war konsistent ein Trend zu einem kleineren Einfluss der Motorik auf die Testleistung gezeigt worden ( $p = .13$  und  $p = .08$ ). Weitere Studien berichteten bereits über kleine bis mittlere Zusammenhänge zwischen Motorikleistungen und der Anzahl bearbeiteter Items in Konzentrationstests (Marschner,

1980; Schmidt-Atzert & Bühner, 1997; Schmidt-Atzert et al., 2006). Es ist davon auszugehen, dass dieser Einfluss bei älteren oder (beispielsweise durch neurologische Erkrankungen) motorisch eingeschränkten Testpersonen größer ausfallen könnte (Taveira & Choi, 2009), daher sollte die Rolle der motorischen Geschwindigkeit für die Testleistung trotz dieses kleineren Effekts nicht unterschätzt werden. Insgesamt sprechen die Ergebnisse früherer Studien sowie die aktuellen Befunde dafür, dass es sich bei der motorischen Geschwindigkeit sehr wahrscheinlich um eine Komponente der Leistung in Konzentrationstests handelt.

### ***8.1.5 Die Rolle des selbstgesteuerten Itemwechsels***

Bisher gibt es keine empirischen Belege dafür, dass die Leistung in Konzentrationstests davon abhängt, wie schnell es einer Testperson gelingt, von einem Item zu einem anderen zu wechseln. In den zwei Studien des ersten Beitrags (Blotenberg & Schmidt-Atzert, 2019c) zeigte sich die Komponente des selbstgesteuerten Itemwechsels mit unterschiedlichen Operationalisierungen (als Geschwindigkeit des Itemwechsels und Itemwechselkosten) als unreliabel und es fand sich kein Zusammenhang mit der Leistung in Konzentrationstests.

Eine ganz wesentliche Frage blieb damit unbeantwortet: Wie wirkt sich die kontinuierliche Bearbeitung eines Tests auf die Informationsverarbeitung aus? Sowohl in theoretischen Überlegungen zur Messung der Konzentrationsfähigkeit als auch in experimentellen Studien zum Einfluss des Präsentationsmodus auf die Validität von Aufmerksamkeits- und Konzentrationstests zeigte sich die kontinuierliche Bearbeitung dieser Tests als ihr kritisches Merkmal (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, et al., 2012). Um dem Rechnung zu tragen, wurde der selbstgesteuerte Itemwechsel als Komponente in das Prozessmodell von Konzentrationstests aufgenommen.

Dabei erwies es sich als äußerst schwierig, diese Prozesskomponente zu messen. Variierte man die Zeitintervalle zwischen den Items, so zeigten sich zwar erwartungskonform starke Effekte auf die Leistung, in dem Sinne dass die Reaktionen deutlich langsamer und fehlerbehafteter waren, wenn die Intervalle sehr kurz waren ( $RSI < 500$  ms) oder gefordert war, dass sofort weitergearbeitet werden musste (siehe auch Jentzsch & Dudschig, 2009; Rabbitt, 1969). Allerdings wirken sich auch noch eine ganze Reihe weiterer, oft variierender Faktoren auf die Dauer der Reaktion aus (individuelle Präferenz für bestimmte Zielreize, Sequenzeffekte vorhergehender Reize, aktuelle Fixation/Sakkade, kurze Ablenkung etc., z. B. Soetens, 1998; Soetens, Boer, & Hueting, 1985). Während sich also der

allgemeinpsychologische Effekt verkürzter Intervalle gut zeigen lässt, sind interindividuelle Unterschiede darin, wie viel Zeit eine Testperson zum Wechsel zwischen den Items braucht, sehr viel schwerer zu erfassen. Beim Wechsel zwischen den Items und damit auch bei der Zeit, die für den Wechsel zwischen Items benötigt wird, handelt es sich womöglich eher um dynamische, flexible Prozesse, die von einer ganzen Reihe von Faktoren abhängen (z.B. Sequenzeffekte, aktuelle Fixation).

Darüber hinaus ist es denkbar, dass sich die kontinuierliche Bearbeitung der Items nicht nur auf einen, sondern auf mehrere, möglicherweise sogar auf alle Prozesskomponenten auswirkt. Wenn Testpersonen gefordert sind, pausenlos an einer Aufgabe zu arbeiten, sollten damit auch erhöhte Anforderungen einhergehen, die einzelnen Teilhandlungen der Aufgabe effizient zu organisieren und auszuführen – und das kontinuierlich. Somit würde im Sinne Dükers (1957) eine höhere Anforderung an die Koordination der Teilhandlungen gestellt. Diese sollte sich auf die Gesamtheit der Prozesskomponenten auswirken und nicht nur isoliert auf den Wechsel zwischen den Items. Damit wäre es zu kurz gegriffen, diese Anforderung in nur einer Prozesskomponente abbilden zu wollen.

#### ***8.1.6 Die Rolle der Vorverarbeitung***

Es konnte im Rahmen der vorliegenden Arbeit gezeigt werden, dass die Leistung in Konzentrationstests auch bedeutsam davon abhängt, wie gut es einer Testperson gelingt, die Vorschau auf nachfolgende Stimuli zu nutzen und diese vorzuverarbeiten. Somit konnte eine neue, stabile Komponente der Leistung in Konzentrationstests demonstriert werden: Die proaktive Vorverarbeitung nachfolgender Stimuli. Dabei war im Vorhinein angenommen worden, dass die Darbietung vieler Stimuli in Konzentrationstests einen Vorteil darstellen könnte, da sie eine Vorschau und mögliche Vorbereitung auf die nachfolgenden Stimuli erlaubt. Ein solcher Vorteil war bereits aus der kognitionspsychologischen Forschung zum Lesen bekannt (Radach & Kennedy, 2013; Rayner, 1975; Schotter et al., 2012) und konnte in der vorliegenden Arbeit auch für Konzentrationstests demonstriert werden. Mittels einer experimentellen Manipulation der Stimulusanordnung konnte gezeigt werden, dass es reliable interindividuelle Unterschiede darin gab, wie gut es einer Testperson gelang, nachfolgende Stimuli vorzuverarbeiten. Das Ausmaß der Vorverarbeitung hing außerdem mit verschiedenen Konzentrationstests (mit unterschiedlichem Testmaterial, unterschiedlichen Aufgaben und Papier-Bleistift- sowie Computerformaten) zusammen. Welche Prozesskomponenten der Konzentrationstestleistung im Einzelnen von der Vorverarbeitung profitieren, kann noch nicht

mit Sicherheit gesagt werden. Studien zum Vorteil der Vorschau beim Lesen zeigen, dass unter anderem orthographische, phonologische, morphologische und wahrscheinlich sogar semantische Eigenschaften des nachfolgenden Wortes vorverarbeitet werden, während das vorherige noch fixiert wird (Schotter et al., 2012). Diese Ergebnisse sind nicht ohne Weiteres auf die Vorverarbeitung in Konzentrationstests übertragbar, lassen aber ahnen, dass auch hier die Verarbeitung nachfolgender Stimuli schon weiter voranschreiten könnte, so dass wahrscheinlich nicht nur reine Wahrnehmungsprozesse, sondern auch schon erste Lösungsprozesse stattfinden.

Wie bereits zu Beginn dieser Arbeit beschrieben, folgen Konzentrationstests keiner gemeinsamen Definition und keinem gemeinsamen, konzeptionellen Rahmen und sind aus der Praxis entstanden (Schmidt-Atzert et al., 2006). Bei der Vorgabe eines Papier-Bleistift-Tests hat die Darbietung vieler Stimuli den großen Vorteil, dass damit eine kontinuierliche, pausenlose Bearbeitung des Tests ermöglicht beziehungsweise abverlangt werden kann. Durch die Vorgabe vieler Stimuli wird aber auch die Vorschau auf die nachfolgenden Stimuli und somit eine Vorbereitung auf nachfolgende Reaktionen möglich. Ein Zusammenhang dieser Leistung mit einigen alltäglichen Konzentrationsleistungen, die von der proaktiven Verarbeitung von Umgebungsreizen profitieren, ist durchaus plausibel. Darunter fällt beispielsweise das Fahren eines Autos im Stadtverkehr oder die Korrektur eines Textes. Es sollte nun geprüft werden, ob sich eine solche Beziehung tatsächlich zeigen lässt, das heißt ob das Ausmaß der Vorverarbeitung in Konzentrationstests mit Außenkriterien der Konzentrationsleistung zusammenhängt. Ein solcher Zusammenhang spräche dafür, dass es sich bei der Vorverarbeitung nachfolgender Stimuli um einen inherenten Bestandteil der Konzentrationsleistung handelt und nicht nur um einen nicht intendierten Nebeneffekt der Testgestaltung. Prinzipiell wäre es für digitale Konzentrationstests ohne weiteres möglich, die Vorschau auf nachfolgende Stimuli zu verhindern. Dafür müsste das nächste Item unmittelbar nach der Reaktion auf ein Vorheriges erscheinen, wie es auch in einigen Bedingungen des vorliegenden Experiments umgesetzt wurde. Ob dies jedoch wünschenswert wäre, müssen weitere Untersuchungen zur Alltagsrelevanz der Vorverarbeitung nachfolgender Stimuli zeigen.

### ***8.1.7 Diskriminante Validität der Prozesskomponenten***

Die Konzentrationsfähigkeit gilt als Voraussetzung für höhere kognitive Fähigkeiten (Lezak, 1995). Diese komplexeren kognitiven Fähigkeiten, wie das schlussfolgernde Denken oder die Arbeitsgedächtniskapazität, sollten neben gemeinsamen Prozessen (wie beispielsweise die Anforderung an effiziente Wahrnehmungsprozesse) zusätzliche, iterative Lösungsprozesse erfordern, die von den postulierten Prozesskomponenten der Leistung in Konzentrationstests nicht adäquat abgebildet werden können (Hunt, 1980; Sternberg, 1977; Unsworth & Engle, 2006, 2007). Erwartungsgemäß zeigte sich im Hinblick auf die diskriminante Validität, dass die postulierten Prozesskomponenten einen kleineren Anteil der Varianz in komplexen kognitiven Aufgaben, wie der Leistung im schlussfolgernden Denken oder der Arbeitsgedächtniskapazität, aufklären konnten. Insgesamt konnte die diskriminante Validität in Bezug auf komplexere kognitive Leistungstests somit bestätigt werden.

Dabei gelang eine klare Abgrenzung des Prozessmodells von der Leistung in Tests der Arbeitsgedächtniskapazität. Diese Tests verlangen, dass Buchstaben, Zahlen oder Figuren im Arbeitsgedächtnis gehalten werden, während andere Aufgaben zu bearbeiten sind (Daneman & Carpenter, 1980; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002; Unsworth & Engle, 2007). Sie erfordern unter anderem die stetige Wiederholung, Aufrechterhaltung und Aktualisierung der Informationen, sowie die kontrollierte Suche und den Abruf der Informationen aus dem Arbeitsgedächtnis (Unsworth & Engle, 2006, 2007). Die Wiedergabe der zu merkenden Buchstaben, Zahlen oder Figuren erfolgt dabei in der Regel ohne Zeitdruck. Daher wurde erwartet, dass die Leistung in diesen Tests nur indirekt von der Wahrnehmungs-, der mentalen Operations- oder der motorischen Geschwindigkeit abhängen sollte. Erwartungsgemäß zeigten sich hier auch nur geringfügige Zusammenhänge.

Für Konzentrationstests und Aufgaben zum schlussfolgernden Denken zeigten sich erwartungskonform Überschneidungen in einigen Prozesskomponenten: So war insbesondere die Wahrnehmungsgeschwindigkeit konsistent ein starker Prädiktor für die Leistung in den Aufgaben zum schlussfolgernden Denken. Ein solcher Einfluss hatte sich bereits in früheren Studien gezeigt (Brand & Deary, 1982; Deary & Stough, 1996; Grudnik & Kranzler, 2001; Kranzler & Jensen, 1989). Hinsichtlich der Lösungsprozesse von Konzentrationstests und Aufgaben zum schlussfolgernden Denken war angenommen worden, dass die simple mentale Operation, die für die Lösung der Items in Konzentrationstests nötig ist, nur einen kleinen Teil der Varianz in den Aufgaben zum schlussfolgernden Denken aufklären könnte. Aufgaben

zum schlussfolgernden Denken verlangen komplexere, mehrstufige Lösungsprozesse, wie beispielsweise den Abgleich und die Aufdeckung der Beziehung zwischen den Wörtern im Itemstamm und die Anwendung dieser Beziehung auf die Antwortoptionen bei verbalen Analogieaufgaben (Sternberg, 1977). Dennoch sagte die Geschwindigkeit einer simplen mentalen Operation in der ersten Studie des ersten Beitrags erfolgreich die Leistung in den Aufgaben zum schlussfolgernden Denken vorher. Dabei ist allerdings zu bedenken, dass die Aufgabe zur Messung der simplen mentalen Operation in dieser ersten Studie relativ komplex war, weil drei Stimuli gleichzeitig relevant waren. Es ist ein gut gesicherter Befund, dass die Anzahl der Stimuli in Wahlreaktionsaufgaben den Zusammenhang mit Intelligenztests erhöht (Jensen, 1982, 1987; Vernon, 1986). Außerdem waren die Konzentrationstests und die Aufgaben zum schlussfolgernden Denken in der ersten Studie ungewöhnlich hoch korreliert. In der zweiten Studie zeigte sich dagegen nur ein Trend für einen kleineren Einfluss der Geschwindigkeit einer simplen mentalen Operation auf die Leistung in den Aufgaben zum schlussfolgernden Denken. Insgesamt konnte die diskriminante Validität des Prozessmodells für Aufgaben zum schlussfolgernden Denken weitgehend bestätigt werden.

#### ***8.1.8 Fazit über die Prozesskomponenten der Leistung in Konzentrationstests***

Zusammengenommen konnten in der vorliegenden Arbeit weitere Erkenntnisse darüber gewonnen werden, welche Prozesskomponenten eine wesentliche Rolle für die Bearbeitung von Konzentrationstests spielen. Konvergierende Evidenz findet sich für einen starken Einfluss der Wahrnehmungsgeschwindigkeit sowie der Geschwindigkeit der Lösungsprozesse auf die Testleistung und eine Tendenz zu einer kleineren Rolle der motorischen Geschwindigkeit. Außerdem gibt es erste Belege dafür, dass die effiziente und erfolgreiche Bearbeitung von Konzentrationstests auch davon abhängt, wie gut es einer Testperson gelingt, nachfolgende Stimuli vorzuverarbeiten. Diese Erkenntnisse sind wichtig, um zu prüfen, ob die gefundenen Prozesskomponenten mit den theoretischen Vorstellungen über das Konstrukt der Konzentrationsfähigkeit korrespondieren. Sie erlauben also einen tieferen Einblick darin, ob Konzentrationstests tatsächlich messen, was sie zu messen vorgeben, als es klassische korrelative Validitätsansätze vermögen. Darüber hinaus kann das Wissen über die Prozesskomponenten auch dazu dienen, ein besseres Verständnis dafür zu entwickeln, wie bestimmte Phänomene in Konzentrationstests entstehen. Ein Beispiel für eines dieser Phänomene ist der relativ große Übungseffekt, dem der folgende Abschnitt gewidmet ist.

## 8.2 Der Übungseffekt in Konzentrationstests

Relativ große Übungseffekte zwischen .37 und bis zu einer Standardabweichung nach einmaliger Testwiederholung sind ein konsistenter Befund vieler Studien, die sich mit dem Einfluss der Übung auf die Leistung in Konzentrationstests befassen haben (Bühner et al., 2006; Scharfen, Peters, et al., 2018; Westhoff & Dewald, 1990). Zudem wächst der Übungseffekt über wiederholte Testungen hinweg weiter an (Westhoff & Dewald, 1990). Über die Ursachen des Übungseffekts ist allerdings noch sehr wenig bekannt (Büttner & Schmidt-Atzert, 2004; Scharfen, Blum, et al., 2018). Im *dritten Beitrag* (Blotenberg & Schmidt-Atzert, 2019b) der vorliegenden Arbeit wurden zum einen die zwei vorhersagestarken Prozesskomponenten der Wahrnehmungs- und mentalen Operationsgeschwindigkeit zugrundegelegt, sowie die motorische Geschwindigkeit betrachtet, für die sich eine Tendenz zu einem kleineren Einfluss auf die Testleistung gezeigt hatte. Zum anderen wurden der selbstgesteuerte Wechsel zwischen den Items und die proaktive Vorverarbeitung nachfolgender Stimuli als weitere potentiell relevante Mechanismen, die von der Übung profitieren könnten, untersucht.

Für die Prozesskomponenten zeigten sich kleinere Übungseffekte in Bezug auf die Wahrnehmungs- und motorische Komponente und ein großer Übungseffekt für die Aufgabe, die neben diesen beiden Komponenten zusätzlich die Lösung des Items verlangte. Somit schienen die Auswirkungen der Übung eher komplex zu sein und gleich mehrere Prozesse zu betreffen. Der größere Anteil am Übungseffekt schien dabei der simplen mentalen Operation zuzukommen, die zur Lösung des Items nötig ist, und möglicherweise auch der Koordination der einzelnen Prozesskomponenten. Zwei weitere betrachtete Mechanismen, der selbstgesteuerte Itemwechsel und die Vorverarbeitung nachfolgender Stimuli, waren durch die Übung nicht effizienter geworden. Dabei galt im Vorhinein insbesondere die Vorverarbeitung nachfolgender Stimuli als vielversprechender Kandidat, weil die Leistung in Konzentrationstests so deutlich von einer effizienten Vorbereitung auf nachfolgende Stimuli profitierte (*siehe Beitrag 2*, Blotenberg & Schmidt-Atzert, 2019a). Allerdings schienen die Versuchspersonen nicht in der Lage zu sein, das Ausmaß des Vorausarbeitens bei wiederholter Testung deutlich auszubauen.

Diese ersten Hinweise darauf, dass sowohl die Wahrnehmungs- und motorische Geschwindigkeit, aber insbesondere die mentale Operation bei der Bearbeitung von Konzentrationstests von der Übung profitiert, sind nicht trivial. Die mentalen Operationen bei



diesen kognitiven Tests sind bewusst kognitiv niederschwellig gewählt, um die Konfundierung mit anderen kognitiven Fähigkeiten und Fertigkeiten möglichst gering zu halten. Es handelt sich hier also um simple mentale Operationen, die danach ausgewählt wurden, dass sie den meisten Menschen von Beginn an leicht fallen. Dennoch scheinen diese über viele Testwiederholungen stark steigerbar zu sein – und das an vielen Stellen im Prozess. Aus diesen komplexen Auswirkungen der Übung lassen sich nur schwer Maßnahmen ableiten, um auf diese Übungseffekte durch eine Abwandlung des Testmaterials zu reagieren. Hätte sich beispielsweise gezeigt, dass der Übungseffekt in Konzentrationstests insbesondere durch eine effizientere Vorverarbeitung nachfolgender Stimuli entsteht, so hätte man diese Tests relativ einfach so abwandeln können, dass keine Vorschau auf nachfolgende Items möglich ist und damit den Übungseffekt deutlich reduzieren können. So bleibt es vorerst bei früheren Empfehlungen zur Reduktion von Übungseffekten, nach denen Instruktionen möglichst gut verständlich und kurz gehalten und wenn möglich bereits vor der Testung vorgegeben werden sollten (eventuell ergänzt um Beispielitems), weil Übungseffekte für Aufgaben mit komplexen Anweisungen besonders groß sind (Powers, 1986; Swinton & Powers, 1983).

Eine Limitation des dritten Beitrags der vorliegenden Arbeit lag darin, dass sich die einzelnen Prozesskomponenten mit der gewählten Methode nur bedingt trennen ließen. Daher wurden gezielt Aufgaben ausgewählt, in denen einzelne Prozesskomponenten schrittweise hinzugenommen wurden. Eine andere Herangehensweise hat sich in den letzten Jahren in der Kognitionpsychologie etabliert: Zur Bestimmung von Prozesskomponenten werden mathematische Modelle kognitiver Prozesse genutzt, darunter das sogenannte Diffusions-Modell für einfache Wahlreaktionsaufgaben (Ratcliff, 1978; für einen Übersichtsartikel siehe auch Ratcliff & McKoon, 2008). Mithilfe dieses Modells wird die Leistung bei der Bearbeitung von Wahlreaktionsaufgaben in ihre Komponenten zerlegt, dafür werden die Reaktionszeitverteilungen richtiger und falscher Antworten für die Bestimmung verschiedener Parameter herangezogen. Dabei ist die Grundannahme des Modells, dass für eine Entscheidung in einer Wahlreaktionsaufgabe kontinuierlich (und in einem rauschigen Prozess) Informationen akkumuliert werden, bis eine Schwelle für eine der zwei Antwortoptionen erreicht ist und eine Reaktion ausgelöst wird. Im Modell werden Parameter für die mittlere Geschwindigkeit und Genauigkeit des Akkumulationsprozesses (sogenannte drift rate,  $v$ ) und für die Streuung (inter-trial variability of drift,  $sv$ ) bestimmt. Außerdem gibt es Parameter für die mittlere Dauer (response time constant,  $t0$ ) und die Streuung (inter-trial

variability of non-decisional components,  $st0$ ) der Prozesse, die nicht zu den Entscheidungsprozessen gehören, also die Enkodierungs- und motorischen Prozesse. Zusätzlich wird ein Parameter für den Antwortstil, also konservativer oder risikofreudiger, ermittelt (threshold separation,  $a$ ). Schließlich gibt es einen Parameter für die Präferenz einer Testperson für die eine oder die andere Antwortoption (sogenannter relative starting point,  $zr$ ) und einen Parameter, der angibt, inwieweit sich diese Präferenz im Verlauf der Wahlreaktionsaufgabe ändert (inter-trial-variability of starting point,  $szt$ ).

Mithilfe des Diffusionsmodells untersuchten Dutilh, Vandekerckhove, Tuerlinckx, und Wagenmakers (2009), wie sich die Übung einer lexikalischen Wahlreaktionsaufgabe (Entscheidung, ob ein Wort real oder fiktiv war) über fünf Testwiederholungen auf die Prozesskomponenten auswirkte. Es zeigte sich, dass zum einen die Geschwindigkeit des Informations-Akkumulationsprozesses profitierte (höhere drift rate  $v$ ), also die Verarbeitung des Stimulus, und zum anderen die entscheidungsunabhängigen Prozesse, also die Enkodierung und die motorische Reaktion (response time constant  $t0$  sinkt). Dabei machten die schnelleren Enkodierungs- und motorischen Prozesse etwa 40 % des Übungsgewinns in den Reaktionszeiten aus. Darüber hinaus wurden die Versuchspersonen über die Testungen hinweg weniger vorsichtig (threshold separation  $a$  sinkt) und die Präferenz für die Antwortoptionen verschob sich, allerdings eher unsystematisch. Insgesamt sprechen die Ergebnisse dieser Untersuchung für komplexe Auswirkungen der Übung auf verschiedene Prozesskomponenten. Dabei beziehen sich die Befunde zwar auf eine lexikalische Wahlreaktionsaufgabe und nicht auf die Bearbeitung eines Konzentrationstests, sind aber durchaus mit den Befunden aus dem Beitrag zu Übungseffekten in Konzentrationstests (*Beitrag 3*, Blotenberg & Schmidt-Atzert, 2019b) kompatibel. In beiden Untersuchungen zeigte sich auf ähnliche Weise, dass sowohl die Lösungsprozesse als auch die Wahrnehmungs- (oder Enkodierungs-) und motorischen Prozesse von der Übung profitierten, wobei der Anteil der letzteren beiden in den Studien konsistent etwas kleiner war. Es lässt sich festhalten, dass es sich bei den Übungseffekten in kognitiven Aufgaben um ein komplexes Phänomen handelt, dass sich an vielen Stellen im Prozess auszuwirken scheint. Zukünftig könnte auch die psychologische Diagnostik von Ansätzen der mathematischen Modellierung kognitiver Prozesse profitieren.

Eine andere Perspektive auf den Übungseffekt stammt aus der kognitionspsychologischen Forschung zum Erlernen von Fertigkeiten. Demnach sind Übungseffekte Ausdruck eines Lernprozesses, in dessen Verlauf Prozesse, die zunächst viel kognitive Kontrolle,

Anstrengung sowie Ressourcen fordern, zunehmend automatisiert ablaufen (Ackerman, 1987; Logan, 1988; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). In der Konsequenz können diese Prozesse immer schneller ausgeführt werden und sind dabei immer weniger fehleranfällig. Es gibt eine Reihe von Theorien dazu, wie diese Automatisierung stattfinden könnte. Diskutiert werden dabei unter anderem unterschiedliche Ausgestaltungen der Grundannahme, dass zu Beginn einer Aufgabe Algorithmen zur Lösungsfindung genutzt werden, während im Laufe der Übung frühere Lösungen direkt aus dem Gedächtnis abgerufen werden können (Logan, 1988, 2002; Nosofsky & Palmeri, 1997; Rickard, 1997) oder die allgemeinere Annahme, dass Aufgabenwiederholungen zu gestärkten Reiz-Reaktions-Assoziationen führen (Ackerman, 1987, 1988; Ackerman & Beier, 2007). Eine wesentliche Voraussetzung für diese Automatisierung ist dabei, dass die Aufgaben konsistent sind, also den immer gleichen Regeln der Informationsverarbeitung folgen und die gleichen Prozesskomponenten beanspruchen (Ackerman, 1987, 1988; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Dies wäre bei Konzentrationstests durchaus der Fall – und tatsächlich werden typische Konzentrationstests auch als relativ einfache Beispiele für das Lernen durch Automatisierung aufgefasst (Ackerman, 1988; Ackerman & Beier, 2007; Hagemester, 2007). Würde Übung tatsächlich zu einer Automatisierung bei der Bearbeitung von Konzentrationstests führen, wie diese Theorien annehmen, so bedeutet das eine Bedrohung für die Validität dieser Tests. Kernannahmen über die Konzentrationsfähigkeit beinhalten, dass beim konzentrierten Arbeiten kontinuierlich Teilhandlungen koordiniert werden müssen, was mit Kontrollanforderungen und mentaler Anstrengung einhergeht. Würden die Prozesse bei der Bearbeitung von Konzentrationstests dagegen im Verlaufe der Übung automatisiert ablaufen, so könnten diese Tests nicht mehr die Geschwindigkeit (und Genauigkeit) beim konzentrierten Arbeiten erfassen (Westhoff & Hagemester, 2001), sondern vornehmlich die Geschwindigkeit automatisierter Prozesse.

Eine solche Automatisierung mit zunehmender Übung sollte sich in sukzessiv geringeren Korrelationen mit höheren kognitiven Leistungen, die hohe Kontrollanforderungen stellen, zeigen. Für Konzentrationstests wurde eine solche potentielle Validitätsveränderung nach Kenntnis der Autorin noch nicht untersucht. Studien zu anderen kognitiven Aufgaben fanden allerdings gemischte Effekte, die nicht konsistent für eine niedrigere g-Faktor-Sättigung der geübten Tests sprechen (Reeve & Lam, 2005; aber siehe auch Te Nijenhuis, van Vianen, & van der Flier, 2007). Beispielsweise untersuchten Albers und Höft (2009), wie sich Übung auf den Zusammenhang zwischen einem mehrfach durchgeführten figuralen (und

konzentrationsähnlichen) Test und einer einmalig durchgeführten Aufgabe zum räumlichen Denken sowie einem einmalig durchgeführten Konzentrationstest auswirkte. Dabei handelte es sich bei dem konzentrationsähnlichen Test um eine computergestützte, figurale Aufgabe, bei der die Testpersonen in selbstbestimmtem Tempo kontinuierlich beantworten mussten, wo sich ein Punkt relativ zu einem Flugzeug befand. Interessanterweise wuchs der Zusammenhang zwischen diesem Test und den anderen Aufgaben im Verlauf der Übung sogar an. Ähnliches zeigte sich auch in den Daten aus Beitrag drei, der sich mit dem Übungseffekt in Konzentrationstests befasste (*Beitrag 3*, Blotenberg & Schmidt-Atzert, 2019b). Hier wurde der Zusammenhang zwischen der Leistung im zweimal vorgegebenen Modell-d2 und der einmalig erhobenen Leistung im schlussfolgernden Denken durch die Übung des Modell-d2 stärker. Dabei lag die Korrelation mit der Leistung im schlussfolgernden Denken bei der ersten Bearbeitung des Modell-d2 mit  $r = .26$  im niedrigen Bereich, bei der zweiten Bearbeitung des Modell-d2 war der Zusammenhang mit  $r = .40$  signifikant größer ( $z = 2.793$ ,  $p = .003$ )<sup>1</sup>. Einschränkung ist dabei zu sagen, dass hier kein klassischer Konzentrationstest eingesetzt wurde, sondern eine modifizierte Variante des Konzentrationstests d2-R, der allerdings die Kriterien konventioneller Konzentrationstests erfüllte und mit dem Original-Test hoch korreliert war ( $r = .68$ ). Somit sprechen bisherige Befunde erst einmal nicht dafür, dass Übung zu deutlich niedrigeren Kontrollanforderungen bei der Bearbeitung dieser konzentrationsähnlichen Tests geführt hat. Zukünftig wäre eine systematische Untersuchung des Zusammenhangs zwischen wiederholt dargebotenen Konzentrationstests und der Leistung im schlussfolgernden Denken wünschenswert, um näher zu bestimmen, ob sich die Validität von Konzentrationstests im Verlaufe der Übung verändert, wie es von einigen Autoren angenommen wird. Falls dies tatsächlich der Fall wäre, so hätte eine potentielle Übung oder ein Training einer Testperson noch drastischere Auswirkungen auf die Interpretierbarkeit der Ergebnisse, als bisher angenommen.

### **8.2.1 Fazit zum Übungseffekt in Konzentrationstests**

Die vorliegende Arbeit liefert erste Hinweise darauf, dass insbesondere die simple mentale Operation zur Lösung des Items durch die Testwiederholung schneller abzulaufen scheint, während Wahrnehmungs- und motorische Prozesse in kleinerem Maße profitieren. Die Befunde sprechen nicht für einen Vorteil der Vorverarbeitung nachfolgender Items oder des selbstgesteuerten Wechsels zwischen Items durch Übung. Insgesamt sind die Effekte der

---

<sup>1</sup> Dabei ging dieser Zusammenhang nicht darauf zurück, dass die Leistung im schlussfolgernden Denken ein guter Prädiktor für die Größe des Übungseffekts im Modell-d2 war. In der vorliegenden Untersuchung waren der Übungseffekt im Modell-d2 und die Leistung im schlussfolgernden Denken unkorreliert ( $r = .08$ ,  $p = .437$ ).

Übung in Konzentrationstests komplexerer Natur. Diese Befunde decken sich mit einer früheren Untersuchung von Übungseffekten in einer lexikalischen Wahlreaktionsaufgabe, in der mithilfe von Diffusionsmodellen gezeigt wurde, dass sowohl die Stimulusverarbeitung als auch entscheidungsunabhängige Prozesse (z.B. Enkodierung und Motorik) profitierten (Dutilh et al., 2009). Schließlich gibt es Theorien aus der Kognitionspsychologie, nach denen die Übung in Konzentrationstests durch eine Automatisierung der Prozesse bei der Bearbeitung von Konzentrationstests entsteht (Ackerman, 1988; Ackerman & Beier, 2007; Hagemester, 2007). Theoretisch passt diese Annahme zu einer früheren Studie, nach der die Koordination von der Übung in Konzentrationstests profitiert (Krumm et al., 2011). Allerdings würde eine automatisierte Bearbeitung von Konzentrationstests eine Bedrohung der Validität dieser Tests darstellen. Erste Untersuchungen zu konzentrationsähnlichen Tests sprechen allerdings nicht dafür, dass Zusammenhänge mit komplexeren kognitiven Tests im Verlaufe der Testwiederholung eines Konzentrationstests abnehmen. Somit scheinen die kognitiven Kontrollanforderungen bei Konzentrationstests durch die Testwiederholung nicht zu sinken, was gegen die Theorie einer Automatisierung spricht. Zukünftige Studien sollten diese Fragestellung dennoch adressieren und systematisch untersuchen.

### **8.3 Limitationen und Stärken**

Es ist wichtig, sich auch die Grenzen der vorliegenden Arbeit bewusst zu machen. So wurde eine sehr homogene Stichprobe erhoben, die zum Großteil aus jungen, weiblichen Psychologiestudentinnen bestand, was Varianz und Repräsentativität der erhobenen Daten eingeschränkt haben dürfte. Beispielsweise könnte, wie bereits zuvor angemerkt, der Einfluss der Motorik auf die Leistung in Konzentrationstests bei einer älteren Stichprobe größer ausfallen (Taveira & Choi, 2009). Außerdem, auch dies wurde bereits angesprochen, ist die Erfassung der postulierten Prozesskomponenten nicht immer isoliert möglich und stellt eine Herausforderung dar (Neubauer & Knorr, 1997; Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015; Townsend, 1990). Schließlich ist insbesondere im ersten Beitrag die Stichprobengröße nur ausreichend, um große Effekte mit der gewünschten statistischen Power im Strukturgleichungsmodell zu entdecken.

Die Stärke der vorliegenden Arbeit liegt darin, dass mittels experimenteller Manipulationen und der Testung von über 300 Versuchspersonen, wesentliche Prozesskomponenten bei der Bearbeitung von Konzentrationstests untersucht werden konnten. Außerdem wurden verschiedene Konzentrationstests eingesetzt, die figurales, numerisches und verbales

Aufgabenmaterial verwendeten und entweder als Papier-Bleistift-Version oder als elektronische Version vorgegeben wurden. Ziel war es, die Gruppe der Konzentrationstests in ihrer Breite möglichst gut abzubilden, um eine möglichst hohe Generalisierbarkeit der Ergebnisse auf andere Konzentrationstests zu gewährleisten.

#### **8.4 Fazit und Ausblick**

Frühere experimentelle Untersuchungen hatten sich bereits mit den theoretischen Vorstellungen über die Konzentrationsfähigkeit auseinandergesetzt und geprüft, inwieweit diese in Konzentrationstests abgebildet werden. Dabei wurde demonstriert, dass Konzentrationstests die kontinuierliche Koordination von Teilhandlungen verlangen und dass es sich dabei um eine Komponente der Leistung in Konzentrationstests handelt (Krumm et al., 2011). Diese Befunde passen zu den theoretischen Vorstellungen über die Konzentrationsfähigkeit, nach denen das konzentrierte Arbeiten energetische Ressourcen und die Koordination von Teilhandlungen erfordert (Westhoff & Hagemeister, 2005).

Darüber, welche Teilhandlungen im Einzelnen zu koordinieren sind, herrscht nun zusätzliche Klarheit: Konzentrationstests verlangen die kontinuierliche Wahrnehmung von Items, simple mentale Operationen, um zur Lösung über die Items zu kommen und motorische Reaktionen, um die Lösung zu indizieren. Außerdem trägt die Vorschau und Vorverarbeitung nachfolgender Items zur Leistung in diesen Tests bei. Erstere Komponenten passen zu den theoretischen Vorstellungen über die Prozesse, die bei der Bearbeitung von Konzentrationstests eine Rolle spielen sollten. Letztere Komponente, also die proaktive Vorverarbeitung nachfolgender Stimuli, sollte noch kritisch darauf geprüft werden, ob sie zum Konstrukt der Konzentrationsfähigkeit gehört.

Die gewonnenen Erkenntnisse können und wurden auch bereits für eine gezieltere Untersuchung des wiederholt beobachteten, relativ großen Übungseffekts genutzt, der Konzentrationsforscher schon seit langem beschäftigt. Hier konnten erste Schlussfolgerungen über potentiell betroffene Prozesskomponenten gezogen werden, es ist aber noch weitere Forschung nötig, um die komplexen Effekte der Übung auf die Informationsverarbeitung in Konzentrationstests zu enträtseln. Die Kenntnis der Prozesskomponenten von Konzentrationstests könnte zukünftig auch dabei helfen, Beeinträchtigungen in spezifischen Bereichen zu erkennen und beispielsweise besser zu verstehen, warum Menschen mit bestimmten neurologischen Erkrankungen oder psychischen Störungen in diesen Tests

schlechter abschneiden (Brickenkamp et al., 2010; Moosbrugger & Oehlschlägel, 2011). Es wäre dabei möglich zu untersuchen, welche Prozesskomponenten im Einzelnen betroffen sind, also verzögert oder möglicherweise nicht richtig ablaufen. Mit diesem Wissen könnten in einem nächsten Schritt spezifische Trainings- und Fördermaßnahmen abgeleitet und entwickelt werden, um evidenzbasierte Behandlungen bei solchen Erkrankungen zu verbessern.

Gleichzeitig ist die Entwicklung und Validierung des Prozessmodells selbst noch längst nicht abgeschlossen. Bisher findet sich konvergierende Evidenz für die Prozesskomponenten der Wahrnehmungsgeschwindigkeit, der Geschwindigkeit einer simplen mentalen Operation und für einen kleineren Einfluss der motorischen Geschwindigkeit. Zusätzlich sprechen erste Ergebnisse für eine neue Komponente der Testleistung, die proaktive Vorverarbeitung nachfolgender Stimuli. Die Rolle der Fokussierungsleistung in Konzentrationstests sollte wegen methodischer Probleme bei der Messung und unzureichender Reliabilität des ermittelten Index erneut untersucht werden. Außerdem ist bislang noch unklar, wie sich die – für Konzentrationstests so charakteristische – fortlaufende und selbstgesteuerte Bearbeitung von Stimuli auf der Prozessebene auswirkt. Um die Mechanismen hinter diesem Phänomen weiter zu untersuchen, wären Ansätze der mathematische Modellierung kognitiver Prozesse vielversprechend. In Folgestudien könnten diese mit den verwendeten experimentellen Validierungsstrategien kombiniert werden. Insgesamt wäre ein noch stärkerer Austausch der psychologischen Diagnostik und der Kognitionpsychologie für beide Seiten fruchtbar.

In Bezug auf die Generalisierbarkeit des Prozessmodells und der bisher untersuchten Komponenten ist zu sagen, dass für die Validierung des Prozessmodells sehr unterschiedliche Konzentrationstests eingesetzt wurden, in denen sowohl figurales, als auch numerisches und verbales Aufgabenmaterial verwendet wurde. Zudem wurde einer der Konzentrationstests als elektronische Version, andere als Papier-Bleistift-Versionen dargeboten. Die postulierten Prozesskomponenten wurden also an sehr verschiedenen Konzentrationstests untersucht, daher sind die gefundenen Ergebnisse mit hoher Wahrscheinlichkeit auch auf andere Konzentrationstests, sowie Tests der Wahrnehmungs-, Verarbeitungs- und Bearbeitungsgeschwindigkeit, sofern sie die Kriterien typischer Konzentrationstests erfüllen, übertragbar (Krumm, Schmidt-Atzert, Michalczyk, et al., 2008; Schmidt-Atzert et al., 2006).

Es bleibt zusammenzufassen, dass die psychologische Diagnostik von experimentellen Ansätzen profitieren und mit ihrer Hilfe Fragestellungen adressieren kann, die im Rahmen

klassischer korrelativer Validitätsstudien kaum prüfbar sind. Eine noch engere Verzahnung von psychometrischen und experimentellen Forschungsansätzen und Theorien wäre wünschenswert. Für die umfassende Validierung eines psychologischen Testverfahrens sollten auch die Prozesse bei der Bearbeitung eines Tests experimentell untersucht und anschließend kritisch darauf geprüft werden, ob diese zu dem Konstrukt passen, das mit dem Test gemessen werden soll. Gelingt es, diese experimentellen Validitätsbelege zu erbringen, so werden weitere stichhaltige Argumente dafür vorgebracht, dass ein Test tatsächlich misst, was er zu messen vorgibt und dass die Schlüsse, die aus einem Testergebnis gezogen werden, fundiert – und valide – sind, ganz im Sinne der neuesten *Standards for Educational and Psychological Assessment* (AERA et al., 2014).



## 9. Literaturverzeichnis

- Abels, D. (1974). *Konzentrations-Verlaufs-Test (KVT)* (2. Auflage). Göttingen: Hogrefe.
- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, 102, 3–27.
- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117, 288–318.
- Ackerman, P. L., & Beier, M. E. (2007). Further explorations of perceptual speed abilities in the context of assessment methods, cognitive abilities, and individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 13, 249–272.
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, 131, 567–589.
- AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Albers, F., & Höft, S. (2009). Übungseffekte bei einem computergestützten Test zum räumlichen Vorstellungsvermögen. *Diagnostica*, 55, 71–83.
- Arnold, W. (1975). *Der Pauli-Test* (5. Auflage). Berlin: Springer.
- Bartenwerfer, H. (1964). Allgemeine Leistungstests. In R. Heiss (Hrsg.), *Handbuch der Psychologie, Band VI, Psychologische Diagnostik* (S. 385–410). Göttingen: Hogrefe.
- Becker, N., Preckel, F., Karbach, J., & Raffel, N. (2014). Die Matrizenkonstruktionsaufgabe. *Diagnostica*, 61, 22–33.
- Beckmann, J., & Strang, H. (1993). Konzentration: Überlegungen zu einem vernachlässigten Konstrukt. In *Aufmerksamkeit und Energetisierung. Facetten von Konzentration und Leistung* (S. 11–32). Göttingen: Hogrefe.
- Berg, D., & Imhof, M. (2001). Aufmerksamkeit und Konzentration. In D. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2. Aufl.). Weinheim: Beltz.
- Blotenberg, I., & Schmidt-Atzert, L. (2019a). *On the characteristics of sustained attention test performance - The role of the preview benefit*. Manuskript eingereicht.
- Blotenberg, I., & Schmidt-Atzert, L. (2019b). *On the locus of the practice effect in sustained attention tests*. Manuskript eingereicht.
- Blotenberg, I., & Schmidt-Atzert, L. (2019c). Towards a process model of sustained attention tests. *Journal of Intelligence*, 7, 3.
- Bonnaire, M. (2014). *Sortieren als Konzentrationsaufgabe: Neuentwicklung und Validierung eines Tests zur Konzentrationsmessung*. Marburg: Philipps-Universität Marburg, Fachbereich Psychologie.
- Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, 23, 532–544.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.

- Bourdon, B. (1895). Observations comparatives sur la reconnaissance, la discrimination et l'association. *Revue Philosophique de la France et de l'Étranger*, 40, 153–185.
- Brand, C. R., & Deary, I. J. (1982). Intelligence and „inspection time“. In H. Eysenck (Hrsg.), *A model of intelligence* (S. 133–148). New York: Springer.
- Brickenkamp, R. (2002). *Test d2: Aufmerksamkeits-Belastungs-Test* (9. überarb.). Göttingen: Hogrefe.
- Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *Test d2 - Revision. Aufmerksamkeits- und Konzentrationstest (d2-R)*. Göttingen: Hogrefe.
- Bühner, M., Ziegler, M., Bohnes, B., & Lauterbach, K. (2006). Übungseffekte in den TAP Untertests Test Go/Nogo und Geteilte Aufmerksamkeit sowie dem Aufmerksamkeits-Belastungstest (d2). *Zeitschrift für Neuropsychologie*, 17, 191–199.
- Büttner, G., & Schmidt-Atzert, L. (2004). *Diagnostik von Konzentration und Aufmerksamkeit*. Göttingen: Hogrefe.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22.
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Daseking, M., & Putz, D. (2015). TBS-TK Rezension: „Test d2-R. Aufmerksamkeits- und Konzentrationstest“. *Psychologische Rundschau*, 66, 265–267.
- Deary, I. J., & Stough, C. (1996). Intelligence and inspection time. *American Psychologist*, 51, 599–608.
- Düker, H. (1957). *Leistungsfähigkeit und Keimdrüsenhormone*. München: J.A. Barth.
- Düker, H., & Lienert, G. A. (1959). *Der Konzentrations-Leistungs-Test (KLT)*. Göttingen: Hogrefe.
- Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E. J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin and Review*, 16, 1026–1036.
- Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179–197.
- Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, 2, 101–118.
- Fleishman, E. A., & Hempel, W. E. (1954). Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 19, 239–252.
- Gilmore, G. C., Royer, F. L., Gruhn, J. J., & Esson, M. J. (2004). Symbol – digit substitution and individual differences in visual search ability. *Intelligence*, 32, 47–64.
- Goldhammer, F., & Moosbrugger, H. (2006). Aufmerksamkeit. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 16–33). Heidelberg: Springer.
- Grudnik, J. L., & Kranzler, J. H. (2001). Meta-analysis of the relationship between intelligence and inspection time. *Intelligence*, 29, 523–535.
- Hagemeister, C. (2007). How useful is the power law of practice for recognizing practice in concentration tests? *European Journal of Psychological Assessment*, 23, 157–165.

- Hagemeister, C., Scholz, A., & Westhoff, K. (2002). Wie kann man Geübtheit in Konzentrationstests erkennbar machen? *Zeitschrift für Personalpsychologie*, 1, 94–102.
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373–385.
- Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 75, 242–259.
- Hunt, E. (1980). Intelligence as an information processing concept. *British Journal of Psychology*, 71, 449–474.
- Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test Form 4*. Göttingen: Hogrefe.
- Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Hrsg.), *A model for intelligence* (S. 93–132). Berlin: Springer.
- Jensen, A. R. (1987). The g beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Hrsg.), *The influence of cognitive psychology on testing* (S. 87–142). Hillsdale, NJ: Erlbaum.
- Jentzsch, I., & Dudschig, C. (2009). Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *Quarterly Journal of Experimental Psychology*, 62, 209–218.
- Kranzler, J. H., & Jensen, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence*, 13, 329–347.
- Krumm, S., Hüffmeier, J., & Lievens, F. (2017). Experimental test validation: Examining the path from test elements to test performance. *European Journal of Psychological Assessment*. Advance online publication.
- Krumm, S., Lipnevich, A. A., Schmidt-Atzert, L., & Bühner, M. (2012). Relational integration as a predictor of academic achievement. *Learning and Individual Differences*, 22, 759–769.
- Krumm, S., Schmidt-Atzert, L., Bracht, M., & Ochs, L. (2011). Coordination as a crucial component of performance on a sustained attention test. *Journal of Individual Differences*, 32, 117–128.
- Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, 37, 347–364.
- Krumm, S., Schmidt-Atzert, L., & Eschert, S. (2008). Investigating the structure of attention: How do test characteristics of paper-pencil sustained attention tests influence their relationship with other attention tests? *European Journal of Psychological Assessment*, 24, 108–116.
- Krumm, S., Schmidt-Atzert, L., Michalczyk, K., & Danthiir, V. (2008). Speeded paper-pencil sustained attention and mental speed tests. *Journal of Individual Differences*, 29, 205–216.
- Krumm, S., Schmidt-Atzert, L., Schmidt, S., Zenses, E. M., & Stenzel, N. (2012). Attention tests in different stimulus presentation modes. *Journal of Individual Differences*, 33, 146–159.
- Lezak, M. D. (1995). *Neuropsychological Assessment* (3rd ed.). New York: Oxford

- University Press.
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, 58, 981–1007.
- Lievens, F., Reeve, C. L., & Heggstad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology*, 92, 1672–1682.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.
- Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, 109, 376–400.
- Marschner, G. (1980). *Revisions-Test*. Göttingen: Hogrefe.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues* (S. 136–181). New York: Guilford Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1–10.
- Meyer-Baron, M., Schaeper, M., & Seeber, A. (2002). A meta-analysis for neurobehavioural results due to occupational mercury exposure. *Archives of Toxicology*, 76, 127–136.
- Mirsky, A. F., Anthony, B. J., Duncan, C. C., Ahearn, M. B., & Kellam, S. G. (1991). Analysis of the elements of attention: A neuropsychological approach. *Neuropsychology Review*, 2, 109–145.
- Mittring, G., & Rost, D. H. (2008). Die verflixten Distraktoren. *Diagnostica*, 54, 193–201.
- Moosbrugger, H., & Goldhammer, F. (2006). Aufmerksamkeits- und Konzentrationsdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 83–102). Heidelberg: Springer.
- Moosbrugger, H., & Oehlschlägel, J. (1996). *Frankfurter Aufmerksamkeits-Inventar (FAIR)*. Göttingen: Hogrefe.
- Moosbrugger, H., & Oehlschlägel, J. (2011). *Frankfurter Aufmerksamkeits-Inventar 2 (FAIR-2)*. Bern: Huber.
- Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference control in children with and without ADHD: A systematic review of Flanker and Simon task performance. *Child Neuropsychology*, 15, 321–342.
- Neubauer, A. C., & Knorr, E. (1997). Elementary cognitive processes in choice reaction time tasks and their correlations with intelligence. *Personality and Individual Differences*, 23, 715–728.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266–300.
- Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2014). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, 47, 1343–1355.
- Oswald, W. D., & Roth, E. (1987). *Zahlen-Verbindungs-Test*. Göttingen: Hogrefe.

- Peper, M. (2004). *Umweltbezogene Verhaltensstörungen. Neuropsychologische Toxikologie der Chlorkohlenwasserstoffe*. Landsberg/Lech: Ecomed Medizin.
- Petermann, F. (2011). Frankfurter Aufmerksamkeits-Inventar 2 (FAIR-2). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 59, 325–326.
- Powers, D. E. (1986). Relations of test item characteristics to test preparation/test practice effects. A quantitative summary. *Psychological Bulletin*, 100, 67–77.
- Rabbitt, P. (1969). Psychological refractory delay and response-stimulus interval duration in serial, choice-response tasks. *Acta Psychologica*, 30, 195–219.
- Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *Quarterly Journal of Experimental Psychology*, 66, 429–452.
- Randall, J. G., & Villado, A. J. (2017). Take two: Sources and deterrents of score change in employment retesting. *Human Resource Management Review*, 27, 536–553.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20, 873–922.
- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7, 65–81.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457–1506.
- Reeve, C. L., & Lam, H. (2005). The psychometric paradox of practice effects due to retesting: Measurement invariance and stable ability estimates in the face of observed score changes. *Intelligence*, 33, 535–549.
- Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, 126, 288–311.
- Royer, F. L. (1971). Spatial orientation and figural information in free recall of visual figures. *Journal of Experimental Psychology*, 91, 326–332.
- Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, 53, 61–97.
- Schaie, K. W. (1989). Perceptual speed in adulthood: Cross-sectional and longitudinal studies. *Psychology and Aging*, 4, 443–453.
- Scharfen, J., Blum, D., & Holling, H. (2018). Response time reduction due to retesting in mental speed tests: A meta-analysis. *Journal of Intelligence*, 6, 6.
- Scharfen, J., Peters, J. M., & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, 67, 44–66.
- Schmidt-Atzert, L., & Amelang, M. (2012). Leistungstests. In *Psychologische Diagnostik*. Berlin: Springer.
- Schmidt-Atzert, L., & Brickenkamp, R. (2017). *Test d2-R - Elektronische Fassung des Aufmerksamkeits- und Konzentrationstests d2-R*. Göttingen: Hogrefe.
- Schmidt-Atzert, L., & Bühner, M. (1997). Was misst der d2-Test? Eine experimentelle Analyse von Testleistungen. *Diagnostica*, 43, 314–326.

- Schmidt-Atzert, L., & Bühner, M. (2000). Aufmerksamkeit und Intelligenz. In K. Schweizer (Hrsg.), *Intelligenz und Kognition: Die kognitiv-biologische Perspektive der Intelligenz* (S. 125–151). Landau: Verlag Empirische Pädagogik.
- Schmidt-Atzert, L., Bühner, M., & Enders, P. (2006). Messen Konzentrationstests Konzentration? Eine Analyse der Komponenten von Konzentrationsleistungen. *Diagnostica*, 52, 33–44.
- Schmidt-Atzert, L., Büttner, G., & Bühner, M. (2004). Theoretische Aspekte von Aufmerksamkeits-/Konzentrationsdiagnostik. In G. Büttner & L. Schmidt-Atzert (Hrsg.), *Diagnostik von Konzentration und Aufmerksamkeit* (S. 3–22). Göttingen: Hogrefe.
- Schmidt-Atzert, L., & Ising, M. (1997). Ein Beitrag zur Konstruktvalidität von d2 und Revisionstest. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18, 241–250.
- Schmidt-Atzert, L., Krumm, S., & Bühner, M. (2008). Aufmerksamkeitsdiagnostik: Ableitung eines Strukturmodells und systematische Einordnung von Tests. *Zeitschrift für Neuropsychologie*, 19, 59–82.
- Schmidt, M., Trueblood, W., Merwin, M., & Durham, R. L. (1994). How much do "attention" tests tell us? *Archives of Clinical Neuropsychology*, 9, 383–394.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66.
- Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, and Psychophysics*, 74, 5–35.
- Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A., & Bergmann, K. (2015). Decomposing the relationship between mental speed and mental abilities. *Intelligence*, 51, 28–46.
- Schweizer, K. (2005). An overview of research into the cognitive basis of intelligence. *Journal of Individual Differences*, 26, 43–51.
- Schweizer, K., & Koch, W. (2003). Perceptual processes and cognitive ability. *Intelligence*, 31, 211–235.
- Schweizer, K., & Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence*, 32, 329–347.
- Schweizer, K., Moosbrugger, H., & Goldhammer, F. (2005). The structure of the relationship between attention and intelligence. *Intelligence*, 33, 589–611.
- Schweizer, K., Zimmermann, P., & Koch, W. (2000). Sustained attention, intelligence, and the crucial role of perceptual processes. *Learning and Individual Differences*, 12, 271–286.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B*, 298, 199–209.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–190.
- Soetens, E. (1998). Localizing sequential effects in serial choice reaction time with the information reduction procedure. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 547–568.
- Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental*

- Psychology: Human Perception and Performance*, 11, 598–616.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, 15, 201–293.
- Steinborn, M. B., Flehmig, H. C., Westhoff, K., & Langner, R. (2008). Predicting school achievement from self-paced continuous performance: Examining the contributions of response speed, accuracy and response speed variability. *Psychology Science*, 50, 613–634.
- Steinborn, M. B., Langner, R., Flehmig, H. C., & Huestegge, L. (2017). Methodology of performance scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines. *Psychological Assessment*, Advance online publication.
- Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, 84, 353–378.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Hrsg.), *Information Processing in Motor Control and Learning* (S. 117–152). New York: Academic Press.
- Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability - And a little bit more. *Intelligence*, 30, 261–288.
- Swinton, S. S., & Powers, D. E. (1983). A study of the effects of special preparation on GRE analytical scores and item types. *Journal of Educational Psychology*, 75, 104–115.
- Taveira, A. D., & Choi, S. D. (2009). Review study of computer input devices and older users. *International Journal of Human-Computer Interaction*, 25, 455–474.
- Te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, 35, 283–300.
- Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, 1, 46–54.
- Unsworth, N., & Engle, R. W. (2006). A temporal-contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language*, 54, 346–362.
- Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133, 1038–1066.
- Vernon, P. A. (1986). The g-loading of intelligence tests and their relationship with reaction times: A comment on Ruchalla et al. *Intelligence*, 10, 93–100.
- Vickers, D., Nettelbeck, T., & Willson, R. J. (1972). Perceptual indices of performance: The measurement of ‘inspection time’ and ‘noise’ in the visual system. *Perception*, 1, 263–295.
- Vickers, D., & Smith, P. L. (1986). The rationale for the inspection time index. *Personality and Individual Differences*, 7, 609–623.
- Westhoff, K. (1989). Übungsabhängigkeit von Leistungen in Konzentrationstests. *Diagnostica*, 35, 122–130.
- Westhoff, K. (1995). Aufmerksamkeit und Konzentration. In M. Amelang (Hrsg.), *Enzyklopädie der Psychologie: Bd. CVIII 2, Verhaltens- und Leistungsunterschiede* (S.

- 375–402). Göttingen: Hogrefe.
- Westhoff, K., & Dewald, D. (1990). Effekte der Übung in der Bearbeitung von Konzentrationstests. *Diagnostica*, 36, 1–15.
- Westhoff, K., & Graubner, J. (2003). Konstruktion eines komplexen Konzentrationstests. *Diagnostica*, 49, 110–119.
- Westhoff, K., & Hagemester, C. (2001). Konzentrationstraining. In *Handbuch kognitives Training* (S. 515–583). Göttingen: Hogrefe.
- Westhoff, K., & Hagemester, C. (2005). *Konzentrationsdiagnostik*. Lengerich: Pabst.
- White, A. P., & Zammarelli, J. E. (1981). Convergence principles: Information in the answer sets of some multiple-choice intelligence tests. *Applied Psychological Measurement*, 5, 21–27.
- White, C. N., Ratcliff, R., & Starns, J. S. (2012). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, 63, 210–238.



### **Beitrag 1:**

Blotenberg, I., & Schmidt-Atzert, L. (2019c). Towards a process model of sustained attention tests. *Journal of Intelligence*, 7(1), 3, doi: 10.3390/jintelligence7010003.



Article

# Towards a Process Model of Sustained Attention Tests

Iris Blotenberg \*  and Lothar Schmidt-Atzert

Department of Psychology, Philipps-University of Marburg, Gutenbergstr. 18, 35032 Marburg, Germany; schmidt-atzert@uni-marburg.de

\* Correspondence: iris.blotenberg@staff.uni-marburg.de

Received: 6 December 2018; Accepted: 24 January 2019; Published: 28 January 2019



**Abstract:** Taking up new approaches and calls for experimental test validation, in the present study we propose and validate a process model of sustained attention tests. Four sub-components were postulated: the perception of an item, a simple mental operation to solve the item, a motor reaction, and the shift to the next item. In two studies, several cognitive tasks and modified versions of the d2-R test of sustained attention were applied in order to determine performance in the proposed sub-components. Their contribution for the prediction of performance in sustained attention tests and tests of higher cognitive abilities was assessed. The sub-components of the process model explained a large amount of variance in sustained attention tests, namely 55–74%. More specifically, perceptual and mental operation speed were the strongest predictors, while there was a trend towards a small influence of motor speed on test performance. The measures of item shifting showed low reliabilities and did not predict test scores. In terms of discriminant validity, results of Study 1 indicated that the postulated sub-components were insufficient to explain a large amount of variance in working memory span tasks, in Study 2 the same was demonstrated for reasoning tasks. Altogether, the present study is the first to disentangle sub-components in sustained attention tests and to determine their role for test performance.

**Keywords:** sustained attention; concentration; process model; experimental test validation; processing speed; reasoning; working memory capacity

## 1. Introduction

Tests of sustained attention (sometimes also referred to as concentration) are among the most popular psychological tests in the domain of cognitive abilities and their validity has been demonstrated in numerous studies [1–5]. Yet, as the field of psychological assessment progresses, calls have been made to take a closer look into the processes that underly test scores and gain a deeper understanding of how these tests scores emerge [6–9]. For example, the new *Standards for Educational and Psychological Testing* request—in their first chapter—that validity evidence should not only be provided in terms of the test’s internal structure or its relation to other measures, but also in terms of the involved response processes [10]. Regarding sustained attention tests, knowledge about the sub-components that drive performance in these tests would not only increase our understanding of sustained attention tests and what they measure, but it could also be the starting point to investigate well-documented but poorly understood phenomena like the large practice effects in sustained attention tests, respectively. Along these lines, the aim of the present study was to take a first step towards a process model of sustained attention tests. In the following, the construct as well as the characteristics of sustained attention tests are outlined and based on that, a process model comprising four sub-components is derived.

### 1.1. Sustained Attention Tests

Sustained attention may be defined as the ability to achieve and maintain a cognitive state that is characterized by a continuous focus on the relevant task and constant mental effort or as Schweizer [11] (p. 46) put it “the long-term allocation of processing resources to a specific task demand”. This ability is necessary for higher cognitive performances and for a large number of activities in everyday and professional life [12]. Not surprisingly, tests of sustained attention are among the most frequently used cognitive ability tests and they are applied in almost every psychological discipline from business to clinical to traffic psychology [3–5].

The first tests designed to measure sustained attention were cancellation and mental arithmetic tests and they even go back to the end of the 19th century [13]. Typically, in letter cancellation tests, rows of letters are presented simultaneously and the test-taker is required to look for target stimuli (for example, the letter “a”) and cross out as many of them as possible until the test is over. Similarly, in mental arithmetic tests, many simple equations are presented at the same time and the participants are required to tick as many of the correct equations and to cross out as many of the incorrect equations as possible within a given time limit. The key performance index in these tests is the number of marked items (speed) or the number of correctly marked items (error-corrected speed) [3]. Modern sustained attention tests follow the tradition of these first tests and still use letter cancellation, mental arithmetic, sorting tasks, and other simple tasks [14].

Although stimuli and tasks may vary between different sustained attention tests, there are two main features that these tests have in common: First, the stimulus material and tasks are simple [14,15] and second, they share a characteristic presentation mode, i.e., several stimuli are presented simultaneously and participants are required to respond and deliberately shift between them until the test is over [15–17]. This so-called self-paced mode constitutes an integral feature of sustained attention tests, because it requires the test-taker to constantly stay on task until the test is over [15–17]. In contrast, other attention tests typically apply single stimuli and use fixed inter-stimulus (ISI) or response stimulus (RSI) intervals (also called force-paced mode). Therefore, self-paced sustained attention tests are considered to require higher mental effort, as the test-takers need to continuously allocate processing resources to the task, while force-paced attention tests allow them to take short breaks between successive items [11,15,17,18].

However, while there is a common understanding of the essential features of sustained attention tests, to date, there is no formal model or systematic examination of the sub-components that may drive performance in these tests. It has also been stated that despite its vital role, sustained attention tends to be neglected in the field of cognitive abilities [18,19]. Therefore, in the present paper, we address this knowledge gap and propose and investigate a process model of sustained attention tests.

### 1.2. A Process Model of Sustained Attention Tests

Imagine a typical sustained attention test that, for example, requires continuous letter or symbol cancellation (e.g., the d2-R [1] or the Frankfurt Attention Inventory [20]) or mental arithmetics (e.g., the Pauli Test [21] or the Revision Test [22]) for a prolonged period of time. First of all, the stimuli need to be perceived and since sustained attention tests present many stimuli at the same time, they impose high perceptual demands [23]. Secondly, the item has to be solved. Although these item-solving processes are not identical for cancellation (e.g., stimulus discrimination and response selection) and mental arithmetic tasks (e.g., mental addition, comparison of solutions, response selection), both share a simplicity and clarity of the cognitive task and a simplicity of the required mental operation. Indeed, previous studies already showed that a variety of sustained attention tests, notwithstanding their differences, correlated highly and loaded on the same factor [2,3]. Thus, we believe it is justified to subsume the different item-solving processes under a common sub-component which we refer to as a simple mental operation. Thirdly, all of these tasks require the execution of some type of motor response (e.g., to tick off, cross out or click on an item) and finally, they demand the deliberate shifting to the next item. Consequently, the test takers’ performance should depend on 1) how quickly the

target is perceived (perceptual speed), 2) how quickly the correct reaction to an item is identified (speed of a simple mental operation), 3) how quickly the motor reaction is carried out (motor speed) and additionally, 4) how much time it takes to move on to the next item (item shifting speed; see also References [1,24]).

Previous studies already investigated some of these proposed sub-components: For example, earlier research demonstrated that the correlation between sustained attention and reasoning measures is partly due to what the authors called shared perceptual and discrimination processes [23]. Moreover, it was shown that motor speed played a role for performance in a variety of sustained attention tests [3,22,25]. Finally, the intentional, self-paced shifting between items has repeatedly been mentioned as an important process in these tests [15–18,26]. The purpose of the present study is to extend these findings by exploring the role of the four proposed sub-components for performance in sustained attention tests. Moreover, while we presume these sub-components to be the main predictors of performance in sustained attention tests, in the following, we argue that tests of higher cognitive abilities should require more and also more complex mental operations beyond the sub-components of our process model of sustained attention tests.

### *1.3. Discriminant Validity: Processes in Higher Cognitive Abilities*

Sustained attention is typically regarded as a prerequisite of higher cognitive performances [12]. These more complex cognitive abilities like reasoning (considered the best indicator of  $g$  [27]) or working memory capacity (purportedly involved in a wide range of complex intellectual abilities, see Reference [28] for a review) involve additional and also more complex sub-processes or as Hunt [29] (p. 457) put it “... tests of intelligence ... seem to require the orchestration of several different functions”. Sternberg [30] identified several component processes in analogical reasoning, involving encoding (perception and storage of attributes), inference (discovering and storage of the relation), mapping (discovering the relation between the first and the second half), application (of the analogy to each answer option), justification (determining the superior answer if none fits perfectly), and finally, a response. Thus, while speeded reasoning tasks also require the fast perception of an item and a quick motor reaction, they additionally demand complex, iterative mental operations not covered by our process model of sustained attention tests.

Similarly, tasks of working memory span demand additional cognitive processes beyond the proposed sub-components. These tasks require the processing of new information and the storage and retention of information (that is no longer present) at the same time [31–33], involving processes like rehearsal, maintenance, memory updating, and controlled memory search [33,34]. Hence, while these tasks of higher cognitive abilities share some of the proposed sub-components, they additionally demand more complex mental operations not included in our process model of sustained attention. We would therefore, in terms of discriminant validity, expect our sub-components to be insufficient to explain a large amount of variance in these tests.

### *1.4. Aims of the Present Research*

The main objective of the present study was to provide and validate the proposed sub-components of sustained attention tests by firstly, assessing the speed in these sub-components and secondly, examining to what extent these sub-components predicted performance in conventional sustained attention tests. The inspection time task [35], which is considered to impose mainly perceptual demands [36–38], was selected as a measure of perceptual speed. To assess the speed of the item-solving process, a modified version of the d2-R test of sustained attention was created that allowed the measurement of reactions to single items. Additionally, the simple reaction time task was applied as a prototypical measure of motor speed [39–41]. Furthermore, in the modified d2, the response-stimulus interval between successive items was varied within subjects to assess item shifting. It was anticipated that shorter RSI, which require the test-taker to immediately shift between items, would impede performance compared to longer intervals [42,43].

We then specified a model in which the sub-components predicted performance in sustained attention tests as well as tests of higher cognitive abilities. It was expected that the speed in the four sub-components would be significantly related to performance in sustained attention tests, that is, the faster the test taker perceived, solved, responded to the item, and shifted towards the next item, the higher the test score. In contrast, in terms of discriminant validity, more complex cognitive abilities like reasoning or working memory capacity should involve additional and also more complex processes not covered by our process model of sustained attention tests [30,33]. Thus, we would expect the proposed sub-components to explain only a limited amount of variance in these tests.

## 2. Study 1

### 2.1. Materials and Methods

#### 2.1.1. Participants

One-hundred-and-three undergraduates of the Philipps University of Marburg gave informed consent prior to taking part in the study, participated voluntarily, and received partial course credit in exchange. The mean age was 23 years ( $SD = 5.2$ , range = 18–58) and the majority was female (68%). Many studied psychology (43%), followed by economics (11%), educational sciences (11%), and other subjects. They had a mean educational training of 3.5 semesters ( $SD = 1.7$ ). The participants reported a normal or corrected-to-normal vision.

#### 2.1.2. Procedure

Participants were tested in groups of two to five in a laboratory. Each test session took about four hours including three ten minutes breaks. First, a pre-experimental questionnaire, the inspection time task, the simple reaction time task, the modified version of the d2 and the Revision Test were applied. After a break, they were followed by the electronic version of the d2-R and three reasoning tasks. The second break was followed by three verbal sustained attention tasks, the simple reaction time task (second administration), the modified d2 (second administration), and three working memory span tasks. After the third and final break and irrelevant for the present study, further cognitive tests and a final questionnaire were administered.

#### 2.1.3. Measures

Several cognitive tasks were applied to test the proposed process model of sustained attention tests. As the overall aim was to gain a better understanding of the processes involved in these tests, we applied three sustained attention tests, namely the d2-R [1], the Revision Test [22], and three verbal subtasks from the Berlin Intelligence Structure (BIS) Test [44] (sustained attention and processing speed, as measured in the BIS, are conceptually similar and have been shown to measure the same construct [3,45]). These tasks were selected based on recommendations by Schmidt-Atzert et al. [3] and covered figural, numerical, and verbal demands [3]. For the tasks that were selected or created to assess the speed in the proposed sub-components of sustained attention tests, the d2-R served as a template. For example, a modified d2 was applied to measure operation speed. Finally, due to their good psychometric properties, figural, numerical and verbal reasoning tasks of the Intelligence-Structure-Test 2000 R [46] and figural, numerical, and verbal complex span tasks [47] were selected as measures of discriminant validity.

#### Conventional Sustained Attention Tests

Test d2-R Electronic Version [1]. In the d2-R, the task was to select the letter “d” with two marks out of “d” s and “p” s with one to four marks. Fourteen computer pages were applied, each consisted of 60 letters and was presented for 20 s (4.40 min in total, plus instructions). The dependent variable was the number of correctly marked items minus the number of confusion errors (error-corrected speed).

Revision Test [22]. In this paper-pencil test, the task was to quickly assess whether a simple equation was correct, to tick the item if it was and to strike it out if it was not. The tests consisted of 15 lines with 44 items and 30 s per line (7.30 min in total, plus instructions). The number of correctly marked equations was assessed.

Berlin Intelligence Structure (BIS) Tasks of Verbal Processing Speed [44]. Three paper-pencil tasks of the Berlin Intelligence Structure Test were administered to measure verbal sustained attention. The first task “classification of words” (CW) was to strike out as many plants as possible from a list of 100 words within 30 s. In the second task “uncompleted words” (UW), 57 incomplete words were presented on a sheet of paper and the missing letters had to be filled in within 50 s. In the third task “part-whole” (PW), a list of 60 words was presented for 40 s and words with a certain semantic relation (a word that was part of another, e.g., month and year) had to be marked. Performance indices were the number of correct responses.

### Tasks Assessing the Sub-Components

**Perceptual Speed: The Inspection Time Task.** In this task, exposure time of a pi-shaped figure with two legs of markedly different lengths was shortened adaptively. Critically, this task was about the shortest exposure time necessary to perceive and correctly indicate which of the two lines was longer, not about reaction times. It is considered a primarily perceptual task and was therefore selected as a measure of perceptual speed [36–38]. The present version of the task was programmed in E-Prime 2.0 [48]. On each trial, a white fixation cross was displayed on a black screen (1000 ms). Then, a white Pi-shaped figure with a shorter (3.5 cm) and a longer (4.5 cm) leg was presented against a black background and afterwards covered by a backward mask (300 ms). Subjects indicated which leg was longer by pressing “c” (left index finger, the key was colored red) or “m” (right index finger, the key was colored green) on a German QWERTZ computer keyboard. Stimulus exposure time was varied using a staircase procedure, i.e., four correct responses led to a shortened exposure time while an incorrect response led to a prolonged exposure time. After three practice trials, the staircase procedure started with an exposure time of 157 ms and decreased or increased in steps of 66 ms (at the beginning) to 16.5 ms (as the experiment proceeded). The experiment ended after 15 reversals (i.e., exposure time which decreased before suddenly increased or vice versa) or 96 trials. The individual inspection time was assessed as the shortest exposure time of the stimulus to which the participant managed to correctly react to for four times in a row.

**Mental Operation and Item Shifting Speed: The Modified d2.** This task was programmed in E-Prime 2.0 [48]. In each trial, three letters were presented simultaneously until the participant made a response. The task was to decide whether there was a d2 among them and press “m” (right index finger, the key was colored green) for a target and “c” (left index finger, the key was colored red) for a nontarget on a German QWERTZ computer keyboard. The number of targets and nontargets was counterbalanced, as was the position of the d2 and the congruence of flanking stimuli (for targets, i.e., whether there were one, two or three d2). Altogether, the task comprised 432 experimental plus 10 practice trials (RSI of 500 ms) and took about 20 minutes. Mean reaction times (RT) served as indicators of mental operation speed<sup>1</sup>. However, the modified d2 also required efficient perceptual and motor processes. Therefore, overlapping cognitive processes between the different indicators of the sub-components were taken into account when specifying structural equation models in the data analysis.

Critically, in the modified d2, the response-stimulus interval (RSI) was varied (blockwise to avoid sequential effects due to mixed RSI [49]) in order to assess item shifting speed. The first experimental block consisted of 18 stimuli and included an RSI of 1000 ms. In the following blocks, the RSI was

<sup>1</sup> Trials after an RSI of 4000 ms were excluded, because the prolonged RSI was introduced to represent an RSI which is so long that it does not further improve performance [49,50].



decreased in steps of 100 ms until it was 0. The last experimental block used an RSI of 4000 ms. From there, the pattern of RSI started all over again. It was expected that short RSI impede performance and that participants' reactions become faster and more accurate as the RSI gets longer [42,43] until the ideal time interval is approached. From there, performance should not benefit from further extension of the RSI [49,50]. Our rationale for determining item shifting speed was that, due to this relationship between RSI and RT, fast reactions should have been preceded by a sufficiently long RSI. Therefore, as an estimator of the time needed for an item shift, the average of RSI preceding the fastest 30% of correct reactions was determined.

**Motor Speed: A Simple Reaction Time (SRT) Task.** Participants were instructed to press the button "c" (Block 1, 20 trials, left index finger, the key was colored red) and afterwards the button "m" (Block 2, 20 trials, right index finger, the key was colored green) as fast as possible as soon as a black dot ( $2 \times 2$  cm) appeared on screen. The dot was presented until a response was made and the next dot appeared after 1000 ms ( $\pm 100$  ms jitter). Altogether, the SRT task consisted of 40 experimental plus 10 practice trials. Mean RT in the SRT task served as indicators of motor speed, because this task is considered a prototypical measure of motor performance [39–41]. However, it also imposes basic perceptual demands which will be taken into account in the data analysis.

#### Divergent Validity: Reasoning and Working Memory Span Tasks

**I-S-T 2000 R Electronic Version [46].** Figural reasoning was assessed using the task "matrices". Participants had to discover the rule underlying the placement of three figures and choose the missing figure out of five response options. In the numerical reasoning task "number series", participants had to identify the underlying rule in a row of seven numbers and fill in the next number in line (time limit for both tasks: 10 min for 20 items). In the verbal reasoning task "verbal analogies", participants had to find the relation between two words and find a word that has a similar relation to the third word out of five response options (time limit: 7 min for 20 items). The dependent variable in each task was the number of correct responses.

**Complex Span Tasks.** Three complex span tasks were applied to assess working memory capacity: Operation span, reading span, and symmetry span [47]. These tasks were similarly structured: They started with a series of practice trials in which 1) a storage task, 2) a processing task, and 3) the combination of the storage and the processing task were practiced. Afterwards, the test trials started in which the storage and the processing tasks were combined. In order to force the participants to focus on the processing task and not only on the recall of the elements, the time for one processing task in the test trials was limited to 2.5 standard deviations of the mean time for a processing task in the practice trials. Additionally, participants received feedback of the percentage of the correctly answered processing tasks. They were informed that their data would only be used if at least 85% of the processing tasks were answered correctly.

In the "symmetry span" task, participants were presented with a set of figures that were either symmetrical or asymmetrical down the vertical axis. After each figure, a  $4 \times 4$  matrix was shown and one square of this matrix was colored red. The position of the red square in the matrix had to be recalled at the end of the set. The numeric version of this group of tasks was an "operation span" task. Participants were presented with a set of arithmetic operations and the task was to judge whether an equation was correct or not. After each arithmetic operation, a letter was presented that had to be recalled at the end of the set. In the "reading span" task, participants were presented with a set of sentences and had to judge whether each sentence was sensible or not. After each sentence, a letter was presented that had to be recalled at the end of the set. Set sizes ranged from 3–5 in the symmetry span task and from 4–6 in the operation and reading spans tasks. Each set size was administered two times (six items comprising 30 symmetry-square, equation-letter or sentence-letter pairs, respectively). The absolute score for each of the three tasks was the number of trials that were recalled in the correct order and without error.

## 2.2. Data Preprocessing

For the RT analysis of the modified d2, errors were excluded and based on the correct trials,  $z$ -values were calculated for each participant in each block; trials with a RT above  $z = 2.5$  and below  $z = -2.5$  were discarded as outliers. The same cut-offs were applied in the SRT task. For the inspection time task, one participant who stated to have pressed the buttons randomly in the post-experimental questionnaire and whose inspection time was above  $z = 4$  was excluded. Regarding the modified d2, more than 30% of incorrect responses led to an exclusion of the participant's data in this task (one participant). For the working memory span task, participants whose accuracy in the processing tasks was below 85% were excluded (see recommendations in Reference [47]; two participants for the verbal, numeric, and figural version).

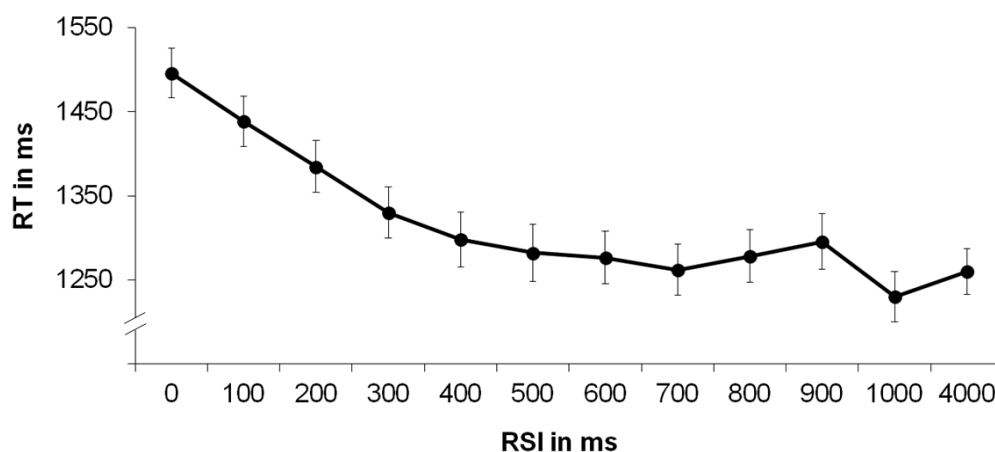
## 2.3. Analysis Strategy

First, as a precondition to assess a potential item shifting component, it was investigated whether short response–stimulus intervals between successive stimuli indeed impeded performance compared to longer response–stimulus intervals as reported in earlier studies [42,43]. Additionally, the descriptive statistics and reliability of the sub-components and cognitive ability tasks were inspected. Finally, to assess the predictive power of the process model, we specified structural equation models in which the four sub-components of perceptual, mental operation, motor and item shifting speed predicted performance in the employed sustained attention tests (convergent validity) and in the reasoning and working memory tasks (discriminant validity).

## 2.4. Results

### 2.4.1. Calculating an Indicator of Item Shifting Speed

The relationship between RSI and RT is depicted in Figure 1. As expected, a repeated-measures ANOVA with the within-subjects factor RSI yielded a significant effect,  $F(7.939, 809.754) = 55.810$ ,  $p < .000$ ,  $\eta^2 p = .354^2$ , indicating increasingly shorter RT as the RSI became longer.



**Figure 1.** Mean reaction time (RT) (in ms) as a function of the duration of the response–stimulus interval (RSI) (in ms).

Post-hoc tests using the Bonferroni correction revealed that increasing the RSI from 0 to 100 ms, from 100 to 200 ms and up until 300 ms elicited a significant decrease in RT, respectively ( $p < .05$ ). These results suggest that the length of the RSI influenced RT and that very short RSI impeded performance.

<sup>2</sup> Mauchly's sphericity test indicated a violation of the sphericity assumption for the factor RSI ( $\chi^2(65) = 210.992$ ,  $p < .001$ ). Therefore, the Greenhouse–Geisser correction was used to adjust degrees of freedom for the main effect of RSI ( $\epsilon = 0.722$ ).



This relationship between RSI and RT was utilized to determine the individual time necessary for an item shift by selecting the fastest 30% of reactions and calculating the average of the preceding RSI.

#### 2.4.2. Descriptive Statistics and Reliability

Means, standard deviations, and reliability estimates of the study measures are presented in Table 1, the correlation matrix is presented in Appendix A. Reliability estimates were high for the modified d2 ( $r_{tt} = .89$  (in retest)–.99 (Cronbach's  $\alpha$ )) but low for the SRT task ( $r_{tt} = .55$  (in retest)–.60 (Cronbach's  $\alpha$ )) and poor for the indicator of item shifting speed ( $r_{tt} = .41$  (in retest)–.51 (Cronbach's  $\alpha$ )). All applied sustained attention tests showed good or excellent reliabilities ( $r_{tt} = .80$  (in retest)–.98 (Cronbach's  $\alpha$ )). Additionally, Cronbach's  $\alpha$  for the reasoning and working memory span tasks ranged from low ( $r_{tt} = .56$  for figural working memory span) to very satisfactory ( $r_{tt} = .90$  for numerical reasoning).

**Table 1.** Means, standard deviations, and reliability estimates of the measures in Study 1.

Tests/Scores	<i>M</i>	<i>SD</i>	<i>r<sub>tt</sub></i>
Speed in the Sub-Components			
Inspection Time (perception)	63.1 <sup>a</sup>	25.4	-
Reaction Time <i>modified d2</i> (simple mental operation)	1321.5 <sup>a</sup>	349.3	.89 <sup>c</sup> /.99 <sup>d</sup>
Simple Reaction Time (motor reaction)	242.8 <sup>a</sup>	30.7	.55 <sup>c</sup> /.60 <sup>d</sup>
Response-Stimulus Interval <i>modified d2</i> (item shifting)	614.5 <sup>a</sup>	50.9	.41 <sup>c</sup> /.51 <sup>d</sup>
Sustained Attention Tests			
Figural (d2-R electronic version)	222.4 <sup>b</sup>	40.2	.98 <sup>d</sup>
Numerical (Revision Test)	388.1 <sup>b</sup>	66.7	.96 <sup>d</sup>
Verbal (BIS UW/CW/PW)	31.0/22.7/12.6 <sup>b</sup>	7.1/6.3/2.7	.96/.93/.80 <sup>d</sup>
Reasoning Tests			
Figural (matrices, I-S-T 2000 R)	10.5 <sup>b</sup>	3.0	.61 <sup>d</sup>
Numerical (number series, I-S-T 2000 R)	13.2 <sup>b</sup>	5.1	.90 <sup>d</sup>
Verbal (verbal analogies, I-S-T 2000 R)	11.5 <sup>b</sup>	2.9	.61 <sup>d</sup>
Working Memory Span Tasks			
Figural (symmetry span)	15.3 <sup>b</sup>	4.2	.56 <sup>d</sup>
Numerical (operation span)	24.0 <sup>b</sup>	5.9	.74 <sup>d</sup>
Verbal (reading span)	21.1 <sup>b</sup>	5.8	.67 <sup>d</sup>

Notes: <sup>a</sup> Average speed in ms, <sup>b</sup> Number of correct items (minus confusion errors for sustained attention tests),

<sup>c</sup> Retest reliability, <sup>d</sup> Cronbach's  $\alpha$ .

#### 2.4.3. Structural Equation Modeling

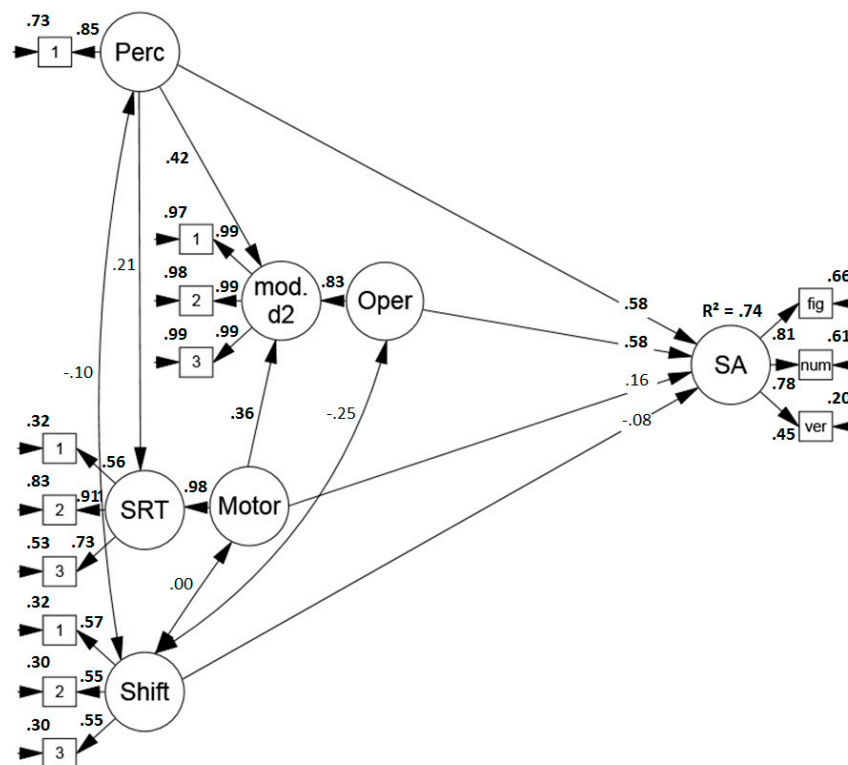
The contribution of the subcomponents of perceptual, mental operation, motor and item shifting speed for the prediction of performance in sustained attention tests, reasoning and working memory span tasks was assessed using Amos 24.0 [51]. Prior to the structural equation modeling, missing values were imputed using the expectation-maximization (EM) algorithm [52].

For the modified d2 (mental operation speed), motor (simple reaction time task), and item shifting speed, trials were randomly assigned to one of three indicators, to assure identifiability of the model [53]. Regarding the adaptive inspection time task, since it cannot be spitted, the reliability estimates for adults as reported in a meta-analysis [54],  $r_{tt} = .73$ , was incorporated in the model via the error term [55]. All study measures were converted so that high values reflected better performance.

In the model, the modified d2 was regressed on perceptual speed (measured as inspection time) and motor speed (measured as simple reaction time) to account for shared perceptual and motor processes and to extract a residual of the modified d2 that would capture mental operation speed. Similarly, as the simple reaction time task, which was applied as a measure of motor speed, also imposed basic perceptual demands, it was regressed on perceptual speed (measured as inspection time) to derive a residual that would represent the sub-component of motor speed. Performance in the applied sustained attention, reasoning and working memory span tests was regressed on the so-obtained sub-components of perceptual, mental operation, motor and item shifting speed.

### Predicting Performance in Sustained Attention Tests

First of all, a model was specified to examine the sub-components' contribution for the prediction of performance in sustained attention tests. The model (see Figure 2) indicated a good fit ( $\chi^2(56) = 58.645$ ,  $p = .379$ , RMSEA = .022 (.000, .066), CFI = .997, SRMR = .049) and the sub-components explained a large amount, namely 74%, of sustained attention test variance. Perceptual and mental operation speed were the main predictors. Descriptively, there was a small influence of motor speed but it failed to reach significance ( $p = .13$ ). The indicator of item shifting speed did not predict performance.

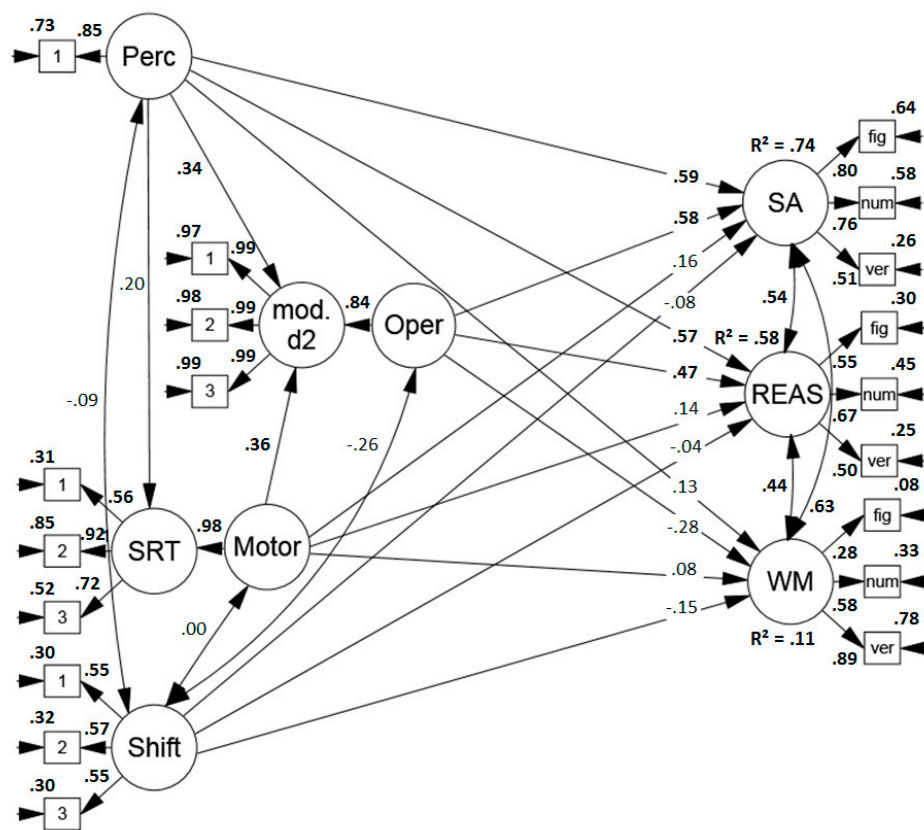


**Figure 2.** Standardized solution of the model predicting performance in sustained attention tests (convergent validity) based on the proposed process model of sustained attention tests. Significant results are printed in boldface. Perc = perceptual speed, mod. d2 = modified d2, Oper. = operation speed, SRT = simple reaction time task, Motor = motor speed, Shift = item shifting speed, SA = sustained attention, fig = figural task, num = numerical task, verb = verbal task.

### Predicting Performance in Sustained Attention, Reasoning and Working Memory Span Tasks

Secondly, the model was extended to assess the predictive power of the sub-components for performance in tests of higher cognitive abilities. The model (see Figure 3) revealed an acceptable to good fit ( $\chi^2(132) = 161.630$ ,  $p = .386^3$ , RMSEA = .047 (.011, .070), CFI = .974, SRMR = .073). Surprisingly, the correlation between sustained attention and reasoning performance was higher than reported in earlier studies (for a meta-analysis, see Reference [56] which reports a mean correlation of  $r = .29$ ). Moreover, unexpectedly, the process model predicted performance in reasoning tests well and explained 58% of the test variance. Not only perceptual speed, but also the speed of the relatively simple mental operation in the modified d2 was a good predictor of reasoning performance. Finally, as anticipated, the four sub-components explained only a modest amount of variance, namely 11% in working memory span performance.

<sup>3</sup> The assumption of multivariate normality could not be confirmed (multivariate kurtosis = 37.441, c.r. = 6.726,  $p < .001$ ). A Bollen–Stine bootstrap procedure (1000 samples) was conducted to obtain a corrected  $p$ -value for the  $\chi^2$ -test.



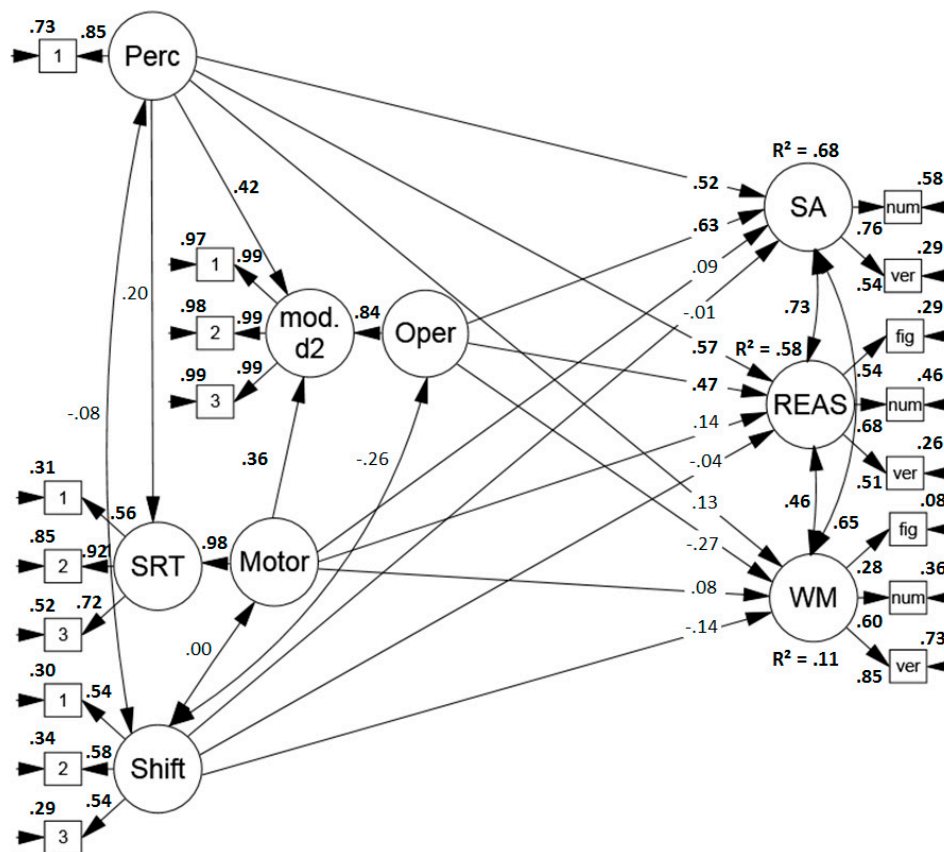
**Figure 3.** Standardized solution of the model predicting performance in sustained attention tests (convergent validity), reasoning, and working memory span tasks (discriminant validity) based on the proposed process model of sustained attention tests. Significant results are printed in boldface. Perc = perceptual speed, mod. d2 = modified d2, Oper = operation speed, SRT = simple reaction time task, Motor = motor speed, Shift = item shifting speed, SA = sustained attention, REAS = reasoning, WM = working memory span, fig = figural task, num = numerical task, ver = verbal task.

## 2.5. Discussion

In Study 1, individual differences in the sub-components explained 74%, and thus, a large amount of variance in sustained attention tests. More specifically, two components, namely perceptual and mental operation speed showed to be strong predictors of test performance, while there was a small, but not significant ( $p = .13$ ) influence of motor speed.

Yet, it is conceivable that these results may partly be due to an overlap between the modified d2 and the d2-R test of sustained attention that served as one of the dependent variables. Therefore, another model was run in which only the Revision Test and the verbal subtasks of the BIS served as indicators of sustained attention test performance (see Figure 4). The model showed a sufficient to good fit ( $\chi^2(115) = 141.459$ ,  $p = .358^4$ , RMSEA = .047 (.005, .072), CFI = .975, SRMR = .071). Altogether, the sub-components explained 68% of the variance in sustained attention performance and again, perceptual and mental operation speed were the strongest predictors. Thus, the explanatory power of the proposed process model for performance in sustained attention tests was not only due to an overlap between the applied tasks.

<sup>4</sup> The assumption of multivariate normality could not be confirmed (multivariate kurtosis = 37.074, c.r. = 6.444,  $p < .001$ ). A Bollen–Stine bootstrap procedure (1000 samples) was conducted to obtain a corrected  $p$ -value for the  $\chi^2$ -test.



**Figure 4.** Standardized solution of the model predicting performance in sustained attention tests (convergent validity, without the d2-R as dependent variable), reasoning and working memory span tasks (discriminant validity) based on the proposed process model of sustained attention tests. Significant results are printed in boldface. Perc = perceptual speed, mod. d2 = modified d2, Oper = operation speed, SRT = simple reaction time task, Motor = motor speed, Shift = item shifting speed, SA = sustained attention, REAS = reasoning, WM = working memory span, fig = figural task, num = numerical task, verb = verbal task.

However, while it was demonstrated that the process model successfully predicted a large amount of variance in sustained attention tests, there was no relationship between item shifting speed and performance in sustained attention tests. On the one hand, this is surprising, as theoretical considerations and earlier research point towards a special significance of the self-paced presentation mode of sustained attention tests [5,15–17], which suggests that a test-taker who succeeds in the fast, deliberate shifting between items should also achieve a higher score in sustained attention tests. On the other hand, we did not succeed in reliably measuring this proposed sub-component with a retest-reliability of .41 and a split-half reliability of .51. We can think of two possible explanations for this: first, item shifting is a flexible process which varies from trial to trial, and thus, represents a state rather than a cognitive ability, or second, our operationalization and measurement of item shifting was inadequate. In order to examine the latter, we took a different approach to investigate item shifting in Study 2.

Moreover, the process model ex hypothesis failed to predict performance in working memory span tasks. However, regarding reasoning performance, not only perceptual speed but also mental operation speed explained a large amount of variance in these tests. While it was not surprising that perceptual processes played a role in reasoning tests (see also Reference [36]), there was the expectation that the simple mental operation required in sustained attention tests would be insufficient to account for the complex processes involved in reasoning. We identified two potential reasons for this finding: First, sustained attention and reasoning tests were unusually highly correlated ( $r = .58$ ) in the present

study, whereas earlier studies reported moderate correlations between these cognitive abilities [3,56]. A reason for this high correlation might have been the relatively late administration (right after several sustained attention tests and more than one hour into the session) of the reasoning tests in the middle of a potentially exhausting, four-hour long test battery of cognitive ability tests. That means, the long test session might have changed the validity of the reasoning tests so that they captured sustained attention abilities especially well. Second, there is much evidence that the complexity of a task and thereby its correlation with reasoning increases with the number of relevant stimuli [57–59]. Therefore, the modified d2 with its three simultaneously relevant stimuli may have been too complex, and thus, might have become a good indicator of reasoning performance.

In Study 2, we address the main issues of Study 1 by (1) operationalizing the proposed item shifting ability differently, (2) presenting the reasoning tests at the beginning of the test battery, and (3) lowering the complexity of the modified d2 in order to measure mental operation speed more adequately.

### 3. Study 2

#### 3.1. Materials and Methods

##### 3.1.1. Participants

One hundred students (72% female) voluntarily participated in Study 2 and received partial course credit in exchange. Their mean age was 22.9 years ( $SD = 4.6$ , range = 18–40) and they had studied 3.2 semesters ( $SD = 2.4$ ) on average in fields like psychology (42%), educational sciences (18%) or economics (9%). Participants gave informed consent in accordance with the Declaration of Helsinki prior to participation.

##### 3.1.2. Procedure

Each test session took about two hours including two ten-minute breaks. It started with a pre-experimental questionnaire, followed by three reasoning tasks. After a break, the inspection time task, the simple reaction time task, the modified d2 and the Revision Test were applied. After another break, the electronic version of the d2-R and the verbal sustained attention tests were administered, followed by a second exposure of the inspection time task, the simple reaction time task, the modified d2 and a final short questionnaire.

##### 3.1.3. Measures

For the assessment of the sub-components of sustained attention tests, another modified d2 was created in order to overcome limitations of the first study. Thus, the assessment of mental operation speed and item shifting differed from Study 1 and is described below. Regarding the assessment of sustained attention and reasoning performance, the same tests were applied as in Study 1.

#### Tasks Assessing the Sub-Components

**Mental Operation Speed and Item Shifting:** The modified d2. Similarly, as in Study 1, the rationale for determining item shifting costs was that responses are faster and more accurate when the length of the RSI is appropriate, whereas too short RSI impede performance [42,43]. In Study 2, item shifting costs were determined as the difference in RT in the conditions with (force-paced) and without an RSI (self-paced).

In the modified version of the d2, the pace of the task was manipulated (force-paced vs. self-paced) and as part of another study, the stimulus arrangement was varied. In the first two blocks, one stimulus was presented at a time and participants had to decide whether the relevant letter was a d2 (response: “m”, right index finger) or not (response: “c”, left index finger). In Blocks 3 and 4, three stimuli were presented at a time, but only the one in the center was relevant. Critically, after the response, there was



an RSI of 500 ms in Blocks 1 and 3 (force-paced conditions) and this RSI was removed in Blocks 2 and 4 (self-paced conditions). Each block consisted of 80 stimuli plus 10 practice trials (15 min in total).

Mean RT of the first two blocks of the modified d2 served as indicators of mental operation speed. As a measure of item shifting costs, the difference in RT of the conditions with (Block 1 and 3) and without an RSI (Block 2 and 4) was assessed. That means, we investigated individual differences in the extent to which RT increased as the participants were required to continuously react to stimuli (self-paced conditions) compared to conditions that included short breaks between successive stimuli (force-paced conditions).

### 3.2. Data Preprocessing

With regard to data cleaning, the same procedure and criteria were applied as in Study 1. For the inspection time task, three participants who mentioned difficulties with the task and whose inspection time was above  $z = 4$  were excluded. For the modified d2, accuracy below 70% in the task led to the exclusion of the participant's data (three participants).

### 3.3. Analysis Strategy

First, as a precondition to determine item shifting costs, it was assessed whether the self-paced mode, which required the deliberate shifting between items, impeded performance compared to the force-paced mode, which allowed short intervals between successive items. Additionally, the descriptive statistics of the study variables were inspected and it was examined whether they showed sufficient reliabilities. Finally, structural equation models were specified in which the four sub-components predicted performance in different sustained attention tests (convergent validity) and reasoning tests (discriminant validity).

### 3.4. Results

#### 3.4.1. Calculating an Indicator of Item Shifting Costs

As a precondition for determining item shifting costs, it was assessed whether RT differed significantly between the force-paced and the self-paced blocks. This was confirmed,  $t(96) = 14.980$ ,  $p < .001$ , indicating longer RT in the self-paced blocks ( $M = 704$  ms,  $SD = 92.5$ ) than in the force-paced blocks ( $M = 628$  ms,  $SD = 86.7$ ). Individual item shifting costs were calculated by first log-transforming response latencies and then subtracting mean RT in the force-paced from mean RT in the self-paced blocks. Log-transformed differences between times equal ratios. We were interested in the individual's relative performance decline or improvement as a function of the self-paced vs. force-paced mode (see also References [32,60]).

#### 3.4.2. Descriptive Statistics and Reliability

Table 2 provides means, standard deviations, and reliability estimates of the study variables. The correlations of the study measures are presented in Appendix B. Reliability was sufficient to high for RT measures of the modified d2 ( $r_{tt} = .68$ – $.89$  (in retest),  $r_{tt} = .92$ – $.97$  (split-half)) and tests of sustained attention ( $r_{tt} = .83$ – $.98$  (Cronbach's alpha)), but low for the inspection time task ( $r_{tt} = .58$  (in retest)), the SRT task ( $r_{tt} = .57$  (in retest)– $.75$  (split-half)) and for item shifting costs ( $r_{tt} = .46$  (in retest)– $.80$  (split-half)). Cronbach's alpha was high for the numerical reasoning test ( $r_{tt} = .83$ ) and considerably lower for the figural ( $r_{tt} = .57$ ) and verbal ( $r_{tt} = .75$ ) reasoning task.

**Table 2.** Means, standard deviations, and reliability estimates of the measures in Study 2.

Tests/Scores	<i>M</i>	<i>SD</i>	<i>r<sub>tt</sub></i>
Modified d2			
Reaction Time <i>force-paced, single stimuli</i>	634.0 <sup>a</sup>	95.5	.68 <sup>c</sup> –.94 <sup>d</sup>
Reaction Time <i>self-paced, single stimuli</i>	702.2 <sup>a</sup>	96.3	.74 <sup>c</sup> –.92 <sup>d</sup>
Reaction Time <i>force-paced, three stimuli</i>	622.6 <sup>a</sup>	85.0	.89 <sup>c</sup> –.95 <sup>d</sup>
Reaction Time <i>self-paced, three stimuli</i>	705.3 <sup>a</sup>	94.0	.85 <sup>c</sup> –.97 <sup>d</sup>
Sub-Components			
Inspection Time (perceptual speed)	53.7 <sup>a</sup>	27.8	.58 <sup>c</sup>
Reaction Time <i>mean single stimuli</i> (simple mental operation)	668.1 <sup>a</sup>	91.2	.74 <sup>c</sup> –.96 <sup>d</sup>
Simple Reaction Time (motor speed)	239.6 <sup>a</sup>	26.3	.57 <sup>c</sup> –.75 <sup>d</sup>
Item shifting <i>reaction time difference (pace)</i>	75.5	49.6	.46 <sup>c</sup> –.80 <sup>d</sup>
Sustained Attention Tests			
Figural (d2-R electronic version)	220.1 <sup>b</sup>	33.9	.98 <sup>d</sup>
Numerical (Revision-Test)	390.0 <sup>b</sup>	74.0	.96 <sup>d</sup>
Verbal (BIS UW/CW/PW)	34.8/25.2/14.9 <sup>b</sup>	8.7/5.7/2.5	.95/.95/.83 <sup>d</sup>
Reasoning Tests			
Figural (matrices, I-S-T 2000R)	11.4 <sup>b</sup>	2.9	.60 <sup>d</sup>
Numerical (number series, I-S-T 2000R)	13.2 <sup>b</sup>	4.2	.83 <sup>d</sup>
Verbal (verbal analogies, I-S-T 2000R)	11.9 <sup>b</sup>	3.1	.67 <sup>d</sup>

Notes: <sup>a</sup> Average speed in ms, <sup>b</sup> Average number of correct items (minus confusion errors for sustained attention tests), <sup>c</sup> Retest reliability, <sup>d</sup> Cronbach's  $\alpha$ .

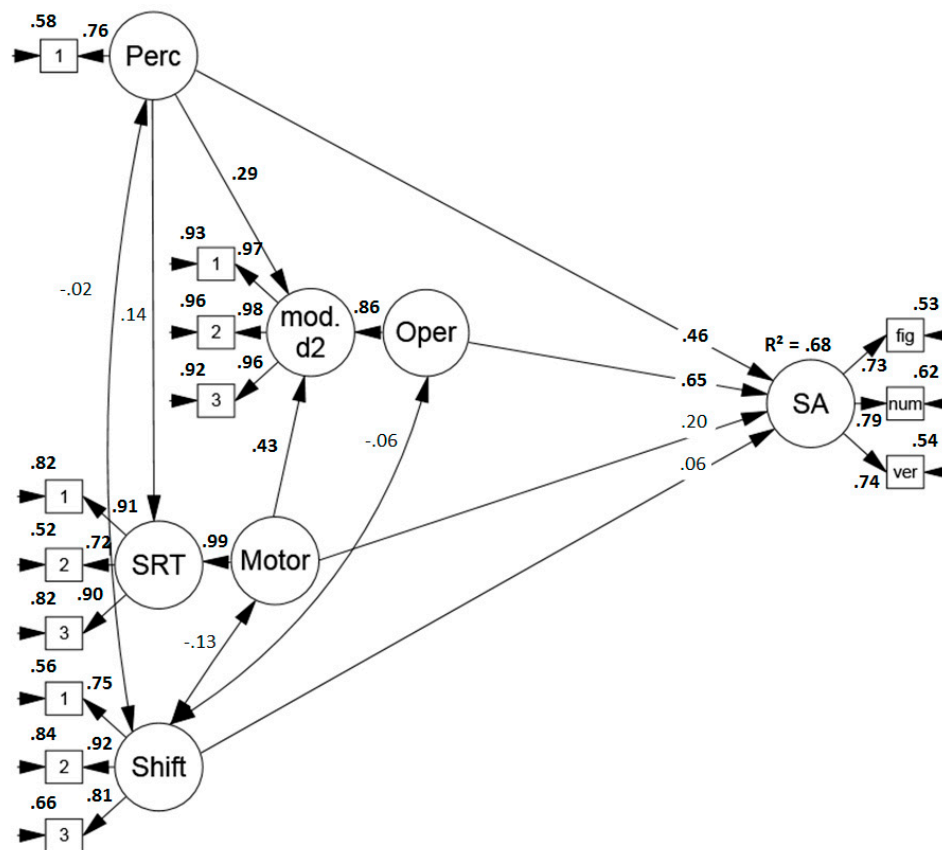
### 3.4.3. Structural Equation Modeling

For the modified d2 (mental operation speed), the simple reaction time task (motor speed), and item shifting, trials were randomly assigned to one of three indicators. Regarding the inspection time task, its retest-reliability in Study 2,  $r_{tt} = .58$ , was incorporated in the model via the error term [55]. All study measures were converted so that high values reflected better performance.

### Predicting Sustained Attention

First, it was assessed to what extent the sub-components predicted performance in the applied sustained attention tests. The model (see Figure 5) revealed a good fit ( $\chi^2(56) = 84.824$ ,  $p = .156^5$ , RMSEA = .072 (.038, .102), CFI = .970, SRMR = .079) and the sub-components predicted 68% of the variance in sustained attention tests. Perceptual and mental operation speed were the strongest predictors and there was a trend towards a minor influence of motor speed ( $p = .08$ ), whereas item shifting costs did not predict test scores.

<sup>5</sup> The assumption of multivariate normality could not be confirmed (multivariate kurtosis = 34.188, c.r. = 8.656,  $p < .001$ ). A Bollen–Stine bootstrap procedure (1000 samples) was conducted to obtain a corrected  $p$ -value for the  $\chi^2$ -test.



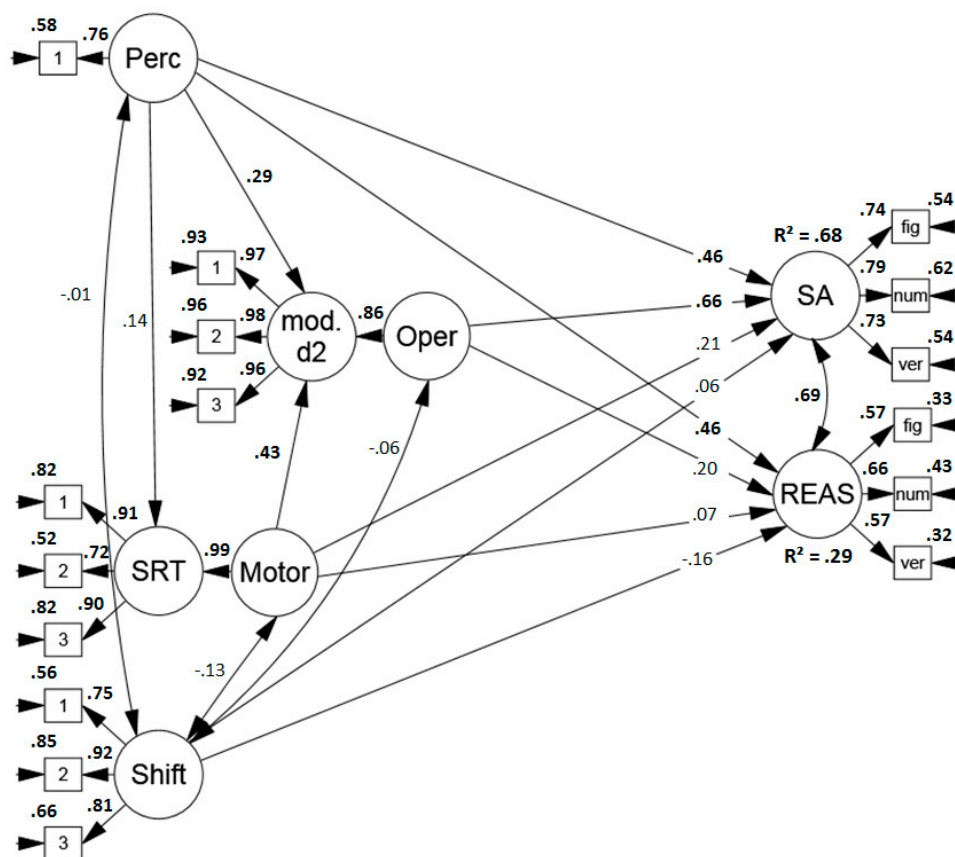
**Figure 5.** Standardized solution of the model predicting performance in sustained attention tests (convergent validity). Significant results are printed in boldface. Perc = perceptual speed, mod.d2 = modified d2, Oper = operation speed, SRT = simple reaction time task, Motor = motor speed, Shift = item shifting speed, SA = sustained attention, fig = figural task, num = numerical task, verb = verbal task.

### Predicting Sustained Attention and Reasoning Test Performance

Second, the model was extended to examine the sub-components' contribution for the prediction of performance in reasoning tasks. The model (see Figure 6) indicated an acceptable to good fit, ( $\chi^2(90) = 124.366$ ,  $p = .197^6$ , RMSEA = .062 (.032, .087), CFI = .967, SRMR = .079). As expected, the process model explained a modest amount of variance in reasoning performance, namely 29%, with perceptual speed being the strongest predictor.

<sup>6</sup> The assumption of multivariate normality could not be confirmed (multivariate kurtosis = 25.676, c.r. = 5.349,  $p < .001$ ). A Bollen–Stine bootstrap procedure (1000 samples) was conducted to obtain a corrected  $p$ -value for the  $\chi^2$ -test.





**Figure 6.** Standardized solution of the model predicting performance in sustained attention tests (convergent validity) and performance in reasoning tests (discriminant validity). Significant results are printed in boldface. Perc = perceptual speed, mod. d2 = modified d2, Oper = operation speed, SRT = simple reaction time task, Motor = motor speed, Shift = item shifting speed, SA = sustained attention, REAS = reasoning, fig = figural task, num = numerical task, ver = verbal task.

### 3.5. Discussion

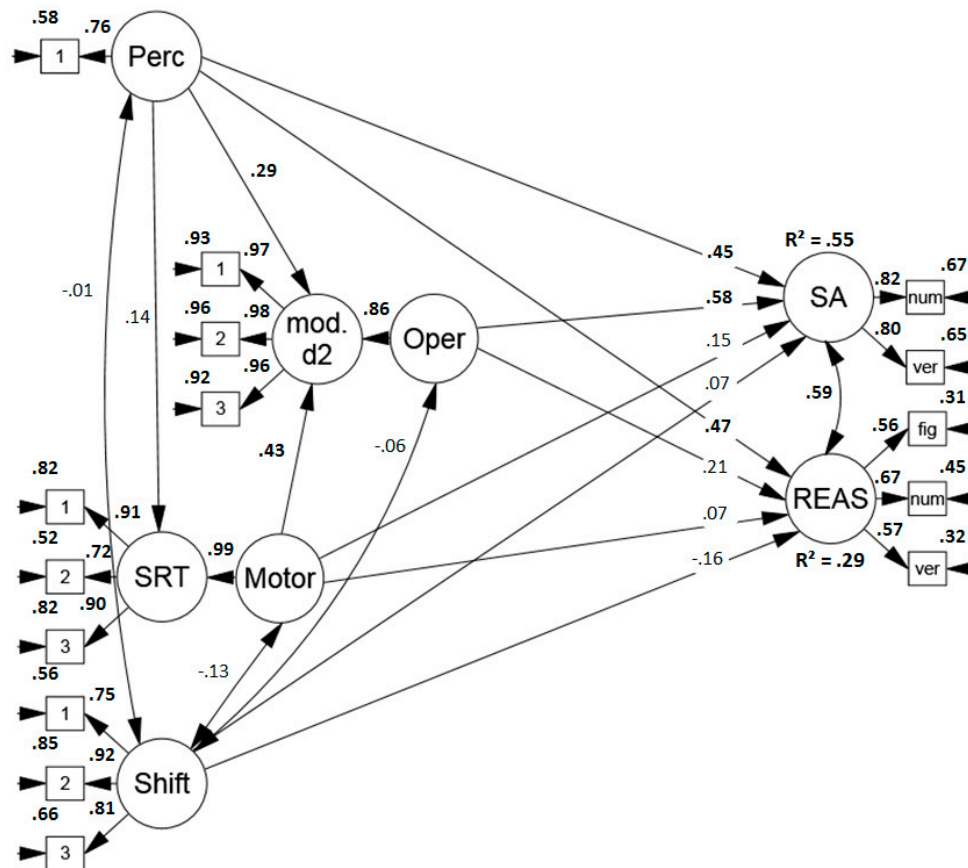
Similarly, as in study 1, the proposed sub-components revealed a high predictive value for performance in SA tests. Overall, the process model explained 68% of the variance in sustained attention tests. Similarly, as in Study 1, perceptual and mental operation speed were the strongest predictors and there was a trend towards a small influence of motor speed ( $p = .08$ ).

However, as the test material of the modified d2 and the d2-R test of sustained attention overlapped considerably, we conducted another analysis and excluded the d2-R test of sustained attention as the dependent variable (see Figure 7). Again, the model fit was good ( $\chi^2(76) = 94.676$ ,  $p = .3497$ , RMSEA = .050 (.000, .080), CFI = .981, SRMR = .070). Altogether, the process model explained less but still a majority, namely 55% of the variance in sustained attention test performance. Still, perceptual and mental operation speed were the strongest predictors.

This time, we took another approach to operationalize item shifting by investigating item shifting costs, that is the decline in performance in a self-paced (without RSI) compared to a force-paced (with an RSI) condition of the modified d2. Although the reliability of the item shifting measure ( $r_{tt} = .46$  (in retest)–.80 (Cronbach's alpha)), was higher than in study 1 ( $r_{tt} = .41$  (in retest)–.51 (split-half)), it again failed to predict performance in sustained attention tests. Note though that RT difference scores are

<sup>7</sup> The assumption of multivariate normality could not be confirmed (multivariate kurtosis = 24.943, c.r. = 5.523,  $p < .001$ ). A Bollen–Stine bootstrap procedure (1000 samples) was conducted to obtain a corrected  $p$ -value for the  $\chi^2$ -test.

typically lower in reliability than mean RT [61]. Overall, results are highly similar to Study 1 with the notable exception that, in accordance with our hypothesis, the process model explained a considerably smaller amount of variance in reasoning performance. In Study 2, only 29% of the variance in reasoning tests could be explained by the proposed process model, thus supporting discriminant validity of the postulated sub-components with regard to reasoning ability.



**Figure 7.** Standardized solution of the model predicting performance in sustained attention tests (convergent validity, without the d2-R as dependent variable) and performance in reasoning tests (discriminant validity). Significant results are printed in boldface. Perc = perceptual speed, mod. d2 = modified d2, Oper = operation speed, SRT = simple reaction time task, Motor = motor speed, Shift = item shifting speed, SA = sustained attention, REAS = reasoning, fig = figural task, num = numerical task, verb = verbal task.

#### 4. General Discussion

In the present study, a generic process model of sustained attention tests was provided based on earlier research on the characteristics of this group of tests. It comprises four essential sub-components: item perception, a simple mental operation, a motor response, and item shifting, hereby taking into account the variety of sustained attention tests. Furthermore, this model was validated by determining the speed in the proposed sub-components and it was demonstrated that two main sub-components, namely perceptual and mental operation speed, successfully predicted a large amount of variance in these tests. Moreover, regarding the third sub-component, motor speed, there was a trend towards a small influence on performance in sustained attention tests. In line with our expectations, the results indicate that while our process model predicted performance in sustained attention tests well, it was by and large insufficient to explain a large amount of variance in tasks of higher cognitive abilities like reasoning or working memory capacity.

#### 4.1. The Role of the Sub-Components for Performance in Sustained Attention Tests

Most importantly, the proposed process model was successful in predicting performance in sustained attention tests. In particular, perceptual and mental operation speed consistently showed to be strong predictors. This is not surprising, as the simultaneous presentation of many stimuli is a typical feature of these tests [16], which makes an efficient perception and processing especially beneficial. Moreover, a fast perception and processing allows a quick shift to the next item or even a preprocessing of the upcoming items [62]. Furthermore, there was a trend towards a small influence of motor speed on performance in both studies (Study 1:  $p = .13$ , Study 2:  $p = .08$ ). Yet, the present sample was young and highly skilled in using computers, which could account for the small variance and low reliability of the measure in both studies [63]. Thus, the role of motor speed could possibly be larger for older participants or for participants who are less familiar with a computer keyboard [64]. Moreover, motor speed was assessed using a computerized task, but only one of the applied sustained attention tests was computerized—the electronic version of the d2-R. Hence, it is conceivable that motor speed as assessed in the current studies could not fully capture the motor processes required in paper–pencil tests. In line with this assumption, descriptively, the influence of motor speed on sustained attention test performance slightly dropped after the electronic d2-R was excluded from the structural equation models.

Unexpectedly, item shifting, assessed as item shifting speed in Study 1 and as item shifting costs in Study 2, did not predict performance in sustained attention tests. Theoretical considerations and earlier research [16,17] pointed towards a special role of the self-paced presentation mode, indicating that the deliberate shifting from one item to another represents a characteristic demand of sustained attention tests. Indeed, in our two studies, whenever the test-takers were required to immediately shift from one item to another, performance was considerably impeded compared to conditions with longer intervals between successive items. However, the retest reliability of the applied item shifting measures was below .50 in both studies. This insufficiently low reliability could either be due to an inappropriate measurement or because item shifting represents a flexible and dynamic process, and thus, a state rather than a trait.

Finally, it is conceivable that the self-paced presentation mode affects several, possibly every sub-component involved when participants work through a test: Indeed, the self-paced mode demands a permanent deployment of mental effort and the enduring organization of the sub-components of sustained attention tests. This organization of sub-components, namely coordination, has repeatedly been suggested as a critical mechanism of sustained attention tests [15,65,66]. Moreover, it has successfully been extracted and its role for test performance has been demonstrated [66]. Therefore, the permanent deployment of mental effort in self-paced sustained attention tests could go along with an enhanced coordination demand which could likely have an impact on several sub-components beyond only the shifting from one item to another.

#### 4.2. Predicting Higher Cognitive Abilities

It was expected that the proposed process model would be insufficient to explain a large amount of variance in more complex cognitive abilities like reasoning or working memory span, because they involve several additional processes [30,33]. In Study 1, it was demonstrated that the process model performed poorly in predicting test scores in working memory span tasks, confirming discriminant validity. This was in line with our expectations, as tests of working memory span measure the ability to store, maintain and retrieve information from working memory [31–33]. Thus, while they also involve perceptual and motor demands, speed in these processes should only be indirectly related to test performance.

Unexpectedly, in Study 1, the sub-components explained a substantial amount of variance in reasoning tests. While it was anticipated that perceptual processes play a role in these tests (see also References [36,54]) we expected the proposed process model to be insufficient to account for the complex mental operations required in reasoning tasks and hence, insufficient to explain a majority

of variance in reasoning performance. We believe that this result is partly due to the comparatively complex task in the modified d2 and due to an unusually high correlation between sustained attention and reasoning tests in Study 1. This high correlation could have been caused by a late presentation of the reasoning tests in a four-hour long test session. Under such circumstances, the validity of the reasoning tests might have shifted from measures of reasoning to good indicators of sustained attention. In Study 2, in accordance with our expectations, the process model explained a comparatively small, but still robust amount of variance in reasoning performance. Again, perceptual speed showed to be a strong predictor of reasoning performance, while the influence of mental operation speed was much smaller. Furthermore, note that the factor loadings of the reasoning tasks on the reasoning factor were relatively low compared to the factor loadings of the sustained attention tests on the sustained attention factor, which led to an increased amount of explained variance for reasoning relative to sustained attention. Altogether, discriminant validity of the process model with regard to higher cognitive abilities could, by and large, be confirmed.

Beyond discriminant validity, the investigation of the relationship between the postulated sub-components and reasoning may also have implications for the study of the processes involved in reasoning performance, i.e., the tasks which are considered the best indicator of *g*. In line with earlier influential research that related individual differences in the speed and efficiency of information processing to *g* [27,36,67–70], we found consistent and substantial associations between perceptual speed and performance in reasoning tasks in both studies. Moreover, in Study 1, using a relatively complex modified d2 (including three simultaneously relevant stimuli), mental operation speed was a good predictor of reasoning performance, while the association between mental operation speed and reasoning was much smaller as a simpler modified d2 (presenting only one stimulus at a time) was applied in Study 2. A possible explanation could be that the more complex mental operations required in the first version of the modified d2 might have more effectively captured the (still much more complex) mental operations that are necessary to solve reasoning tasks (see also References [57–59]). However, note that, as discussed above, measures of reasoning and sustained attention were also unexpectedly highly correlated in that study.

#### 4.3. Limitations

As mentioned earlier, the current samples were not representative with regard to education, age, and sex, as the samples consisted of highly educated, mostly female and young students. Thus, the results cannot be generalized to other samples; the same applies to other tests and tasks. With regard to sample size, a statistical power analysis with simulated data was conducted in Mplus [71]. For both studies, an 80% power to detect significant effects was given for relatively large effects (above .40). Therefore, we should not draw too strong a conclusion about the statistical significance of smaller or moderate effects. However, note that, beyond the significance level, the size of the effect as well as the amount of variance explained in the respective cognitive ability was of special interest in the current study.

Moreover, the correlation between sustained attention and reasoning tests was higher in the present samples than in earlier studies, which report moderate correlations (for a meta-analysis, see Reference [56] which reports a correlation of  $r = .29$ ). Sustained attention tests have been shown to correlate higher with reasoning than other attention tests, which is interpreted in terms of higher cognitive demands due to a permanent mental effort [11,17]. Nevertheless, even with regard to other studies investigating the relationship between sustained attention and reasoning tests, the observed correlation was above comparative studies and samples. This could be attributed to the length of the test sessions, especially regarding the four-hour test session in Study 1. Anyhow, test sessions in the reference studies were shorter but still several hours long, and thus, the observed result can only partly be attributed to the high endurance demand imposed by the long test session.

Finally, our approach of investigating individual differences in the postulated sub-components of sustained attention tests involved measures of perceptual, mental operation, motor and item shifting

speed. However, these sub-components are not necessarily independent or strictly serial [72–74] and even relatively simple cognitive tasks include several of them [75]. Therefore, we specified regression models in order to account for shared variance between tasks with overlapping processes. Nevertheless, even using this approach, we might have not fully disentangled the shared processes between the indicators of the postulated sub-components.

## 5. Conclusions

The present study is, to the best of our knowledge, the first to propose and test a process model of sustained attention tests. In two studies, the convergent validity of the process model was demonstrated, while its discriminant validity with regard to higher cognitive abilities was by and large confirmed. It was demonstrated that two main sub-components, namely perceptual and mental operation speed explained a large amount of variance in sustained attention tests. Moreover, in both studies, there was a trend towards a small influence of motor speed on performance. However, a proposed sub-component of item shifting, which was introduced into the model in order to take the self-paced mode of these tests into account, could not be shown. Thus, for now, the impact of the self-paced mode on information processing in cognitive ability tasks remains to be elucidated. Altogether, the present paper may have taken an important first step towards a deeper understanding of the sub-components that drive sustained attention performance.

**Author Contributions:** I.B. and L.S.-A. conceived and designed the studies, I.B. programmed the tasks, analyzed the data and wrote the manuscript, L.S.-A. reviewed and edited the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

Table A1. Correlation matrix of the measures in Study 1.

Tasks	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Sub-Components														
1. Inspection Time Task	.35 **	.19	−.08	.45 **	.35 **	.17	.22 *	.12	.34 **	.35 **	.14	.24 *	.14	.09
2. Reaction Times <i>modified d2</i>		.42 *	−.24 **	.65 **	.64 **	.27 **	.25 *	.27 **	.35 **	.47 **	.37 **	.20	.04	−.14
3. Simple Reaction Time			−.05	.26 *	.18	.12	.20 *	.12	.26 *	.17	.19	.17	.04	.05
4. Item Shifting Speed				−.30 **	−.14 **	−.03	−.07	−.18 *	−.02	−.11	−.05	−.01	−.00	.08
Sustained Attention														
5. d2-R (electronic version)					.62 **	.21 *	.32 **	.33 **	.37 **	.43 **	.27 **	.29 **	.17	.18
6. Revision Test						.24 *	.35 **	.42 **	.27 **	.50 **	.25 *	.17	.17	.08
7. BIS CW							.57 **	.42 **	.28 **	.11	.39 **	.06	.16	.27 **
8. BIS PW								.47 **	.40 **	.41 **	.36 **	.23 *	.22 *	.32 **
9. BIS UW									.22 *	.20 *	.26 **	.11	.06	.20 *
Reasoning														
10. I-S-T 2000 R Matrices										.37 **	.31 **	.08	.17	.13
11. I-S-T 2000 R Number Series											.32 **	.23 *	.14	.07
12. I-S-T 2000 R Analogies												.25 *	.18	.19
Working Memory Span														
13. Figural Span													.24 *	.22 *
14. Numerical Span														.48 **
15. Verbal Span														

Notes:  $N = 100\text{--}103$ , due to excluded data in some tasks (pairwise deletion). \*  $p < .05$ , \*\*  $p < .01$ .

## Appendix B

Table A2. Correlation matrix of the measures in Study 2.

Tasks	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
Sub-Components											
1. Inspection Time Task	.26 *	.11	.01	.26 *	.34 **	.20	.24 *	.17	.17	.28 **	.16
2. Reaction Time <i>modified d2</i>		.44 **	−.12	.68 **	.56 **	.43 **	.42 **	.39 **	.04	.43 **	.08
3. Simple Reaction Time			−.12	.28 **	.25 *	.10	.04	.17	.01	.13	.20 *
4. Item Shifting Costs				.00	−.02	−.10	.07	.04	−.08	−.14	−.02
Sustained Attention											
5. d2-R (electronic version)					.57 **	.41 **	.26 **	.41 **	.23 *	.39 **	.21 *
6. Revision Test						.56 **	.47 **	.57 **	.25 *	.45 **	.28 **
7. BIS CW							.51 **	.62 **	.23 *	.36 **	.30 **
8. BIS PW								.38 **	.10	.24 *	.10
9. BIS UW									.18	.35 **	.32 **
Reasoning											
10. I-S-T 2000 R Matrices										.35 **	.44 **
11. I-S-T 2000 R Number Series											.33 **
12. I-S-T 2000 R Analogies											

Notes:  $N = 94\text{--}100$ , due to excluded data in some tasks (pairwise deletion). \*  $p < .05$ , \*\*  $p < .01$ .



## References

- Schmidt-Atzert, L.; Brickenkamp, R. *Test d2-R—Elektronische Fassung des Aufmerksamkeits- und Konzentrationstests d2-R*; Hogrefe: Göttingen, Germany, 2017.
- Mirsky, A.F.; Anthony, B.J.; Duncan, C.C.; Ahearn, M.B.; Kellam, S.G. Analysis of the elements of attention: A neuropsychological approach. *Neuropsychol. Rev.* **1991**, *2*, 109–145. [[CrossRef](#)] [[PubMed](#)]
- Schmidt-Atzert, L.; Bühner, M.; Enders, P. Messen Konzentrationstests Konzentration? Eine Analyse der Komponenten von Konzentrationsleistungen. *Diagnostica* **2006**, *52*, 33–44. [[CrossRef](#)]
- Moosbrugger, H.; Goldhammer, F. Aufmerksamkeits- und Konzentrationsdiagnostik. In *Leistung und Leistungsdiagnostik*; Schweizer, K., Ed.; Springer: Heidelberg, Germany, 2006; pp. 83–102.
- Westhoff, K.; Kluck, M.L. Ansätze einer Theorie konzentrativer Leistungen. *Diagnostica* **1984**, *30*, 167–183.
- Bornstein, R.F. Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychol. Assess.* **2011**, *23*, 532–544. [[CrossRef](#)] [[PubMed](#)]
- Borsboom, D.; Mellenbergh, G.J.; Van Heerden, J. The concept of validity. *Psychol. Rev.* **2004**, *111*, 1061–1071. [[CrossRef](#)] [[PubMed](#)]
- Krumm, S.; Hüffmeier, J.; Lievens, F. Experimental test validation: Examining the path from test elements to test performance. *Eur. J. Psychol. Assessment. Adv. Online Publ.* **2017**. [[CrossRef](#)]
- Embretson, S.E. Construct validity: Construct representation versus nomothetic span. *Psychol. Bull.* **1983**, *93*, 179–197.
- American Educational Research Association; American Psychological Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*; American Educational Research Association: Washington, DC, USA, 2014.
- Schweizer, K. An overview of research into the cognitive basis of intelligence. *J. Individ. Differ.* **2005**, *26*, 43–51. [[CrossRef](#)]
- Lezak, M.D. *Neuropsychological Assessment*, 3rd ed.; Oxford University Press: New York, NY, USA, 1995.
- Bartenwerfer, H. Allgemeine Leistungstests. In *Handbuch der Psychologie, Band VI, Psychologische Diagnostik*; Heiss, R., Ed.; Hogrefe: Göttingen, Germany, 1964; pp. 385–410.
- Büttner, G.; Schmidt-Atzert, L. *Diagnostik von Konzentration und Aufmerksamkeit*; Hogrefe: Göttingen, Germany, 2004.
- Westhoff, K.; Hagemester, C. *Konzentrationsdiagnostik*; Pabst: Lengerich, Germany, 2005.
- Krumm, S.; Schmidt-Atzert, L.; Eschert, S. Investigating the structure of attention: How do test characteristics of paper-pencil sustained attention tests influence their relationship with other attention tests? *Eur. J. Psychol. Assess.* **2008**, *24*, 108–116. [[CrossRef](#)]
- Krumm, S.; Schmidt-Atzert, L.; Schmidt, S.; Zenses, E.M.; Stenzel, N. Attention tests in different stimulus presentation modes. *J. Individ. Differ.* **2012**, *33*, 146–159. [[CrossRef](#)]
- Schweizer, K.; Moosbrugger, H. Attention and working memory as predictors of intelligence. *Intelligence* **2004**, *32*, 329–347. [[CrossRef](#)]
- Beckmann, J.; Strang, H. Konzentration: Überlegungen zu einem vernachlässigten Konstrukt. In *Aufmerksamkeit und Energetisierung. Facetten von Konzentration und Leistung*; Hogrefe: Göttingen, Germany, 1993; pp. 11–32.
- Moosbrugger, H.; Oehlschlägel, J. *Frankfurter Aufmerksamkeits-Inventar (FAIR)*; Hogrefe: Göttingen, Germany, 1996.
- Arnold, W. *Der Pauli-Test*; 5. Auflage; Springer: Berlin, Germany, 1975.
- Marschner, G. *Revisions-Test*; Hogrefe: Göttingen, Germany, 1980.
- Schweizer, K.; Zimmermann, P.; Koch, W. Sustained attention, intelligence, and the crucial role of perceptual processes. *Learn. Individ. Differ.* **2000**, *12*, 271–286. [[CrossRef](#)]
- Sanders, A.F. Towards a model of stress and human performance. *Acta Psychol.* **1983**, *53*, 61–97. [[CrossRef](#)]
- Schmidt-Atzert, L.; Bühner, M. Was misst der d2-Test? Eine experimentelle Analyse von Testleistungen. *Diagnostica* **1997**, *43*, 314–326.
- Van Breukelen, G.J.P.; Roskam, E.E.C.I.; Eling, P.A.T.M.; Jansen, R.W.T.L.; Souren, D.A.P.B.; Ickenroth, J.G.M. A model and diagnostic measures for response time series on tests of concentration: Historical background, conceptual framework, and some applications. *Brain Cognit.* **1995**, 147–179. [[CrossRef](#)]



27. Carroll, J.B. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*; Cambridge University Press: New York, NY, USA, 1993.
28. Conway, A.R.A.; Kane, M.J.; Bunting, M.F.; Hambrick, D.Z.; Wilhelm, O.; Engle, R.W. Working memory span tasks: A methodological review and user's guide. *Psychon. Bull. Rev.* **2005**, *12*, 769–786. [[CrossRef](#)] [[PubMed](#)]
29. Hunt, E. Intelligence as an information processing concept. *Br. J. Psychol.* **1980**, *71*, 449–474. [[CrossRef](#)] [[PubMed](#)]
30. Sternberg, R.J. Component processes in analogical reasoning. *Psychol. Rev.* **1977**, *84*, 353–378. [[CrossRef](#)]
31. Daneman, M.; Carpenter, P. Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* **1980**, *19*, 450–466. [[CrossRef](#)]
32. Oberauer, K.; Süß, H.; Wilhelm, O.; Wittman, W. The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence* **2003**, *31*, 167–193. [[CrossRef](#)]
33. Unsworth, N.; Engle, R.W. On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychol. Bull.* **2007**, *133*, 1038–1066. [[CrossRef](#)] [[PubMed](#)]
34. Unsworth, N.; Engle, R.W. A temporal-contextual retrieval account of complex span: An analysis of errors. *J. Mem. Lang.* **2006**, *54*, 346–362. [[CrossRef](#)]
35. Vickers, D.; Smith, P.L. The rationale for the inspection time index. *Pers. Individ. Dif.* **1986**, *7*, 609–623. [[CrossRef](#)]
36. Brand, C.R.; Deary, I.J. Intelligence and “Inspection Time.”. In *A Model of Intelligence*; Eysenck, H., Ed.; Springer: New York, NY, USA, 1982; pp. 133–148.
37. Schweizer, K.; Koch, W. Perceptual processes and cognitive ability. *Intelligence* **2003**, *31*, 211–235. [[CrossRef](#)]
38. Vickers, D.; Nettelbeck, T.; Willson, R.J. Perceptual indices of performance: The measurement of ‘Inspection Time’ and ‘Noise’ in the visual system. *Perception* **1972**, *1*, 263–295. [[CrossRef](#)] [[PubMed](#)]
39. Ackerman, P.L. Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *J. Exp. Psychol. Gen.* **1988**, *117*, 288–318. [[CrossRef](#)]
40. Fleishman, E.A.; Hempel, W.E. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika* **1954**, *19*, 239–252. [[CrossRef](#)]
41. Sternberg, S.; Monsell, S.; Knoll, R.L.; Wright, C.E. The Latency and Duration of Rapid Movement Sequences: Comparisons of Speech and Typewriting. In *Information Processing in Motor Control and Learning*; Stelmach, G.E., Ed.; Academic Press: New York, NY, USA, 1978; pp. 117–152.
42. Jentzsch, I.; Dudschig, C. Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *Q. J. Exp. Psychol.* **2009**, *62*, 209–218. [[CrossRef](#)]
43. Rabbitt, P. Psychological refractory delay and response-stimulus interval duration in serial, choice-response tasks. *Acta Psychol.* **1969**, *30*, 195–219. [[CrossRef](#)]
44. Jäger, A.O.; Süß, H.M.; Beauducel, A. *Berliner Intelligenzstruktur-Test (BIS)*; Hogrefe: Göttingen, Germany, 1997.
45. Krumm, S.; Schmidt-Atzert, L.; Michalczyk, K.; Danthiir, V. Speeded paper-pencil sustained attention and mental speed tests. *J. Individ. Differ.* **2008**, *29*, 205–216. [[CrossRef](#)]
46. Liepmann, D.; Beauducel, A.; Brocke, B.; Amthauer, R. *Intelligenz-Struktur-Test 2000 R*; Hogrefe: Göttingen, Germany, 2007.
47. Oswald, F.L.; McAbee, S.T.; Redick, T.S.; Hambrick, D.Z. The development of a short domain-general measure of working memory capacity. *Behav. Res. Methods* **2014**, *47*, 1343–1355. [[CrossRef](#)] [[PubMed](#)]
48. Schneider, W.; Eschman, A.; Zuccolotto, A. *E-Prime 2.0 Software*; Psychology Software Tools Inc.: Pittsburgh, PA, USA, 2002.
49. Niemi, P.; Näätänen, R. Foreperiod and simple reaction time. *Psychol. Bull.* **1981**, *89*, 133–162. [[CrossRef](#)]
50. Frowein, H.W.; Sanders, A.F. Effects of visual stimulus degradation, S-R compatibility, and foreperiod duration on choice reaction time and movement time. *Bull. Psychon. Soc.* **1978**, *12*, 106–108. [[CrossRef](#)]
51. Airbuckle, J.L. *Amos (Version 24.0)*; IBM SPSS: Chicago, IL, USA, 2016.
52. Allison, P.D. *Missing Data*; Sage Publications: Thousand Oaks, CA, USA, 2001.
53. Anderson, J.; Gerbing, D. Structural equation modeling in practice: A review and recommended two-step approach. *Psychol. Bull.* **1988**, *103*, 411–423. [[CrossRef](#)]
54. Kranzler, J.H.; Jensen, A.R. Inspection time and intelligence: A meta-analysis. *Intelligence* **1989**, *13*, 329–347. [[CrossRef](#)]

55. Brown, T.A. *Confirmatory Factor Analysis for Applied Research*; The Guilford Press: New York, NY, USA, 2006.
56. Schmidt-Atzert, L.; Bühner, M. Aufmerksamkeit und Intelligenz. In *Intelligenz und Kognition: Die kognitiv-biologische Perspektive der Intelligenz*; Schweizer, K., Ed.; Verlag Empirische Pädagogik: Landau, Germany, 2000; pp. 125–151.
57. Jensen, A.R. Reaction Time and Psychometric g. In *A Model for Intelligence*; Eysenck, H.J., Ed.; Springer: Berlin, Germany, 1982; pp. 93–132.
58. Jensen, A.R. The g beyond Factor Analysis. In *The Influence of Cognitive Psychology on Testing*; Ronning, R.R., Glover, J.A., Conoley, J.C., Witt, J.C., Eds.; Erlbaum: Hillsdale, NJ, USA, 1987; pp. 87–142.
59. Vernon, P.A. The g-loading of intelligence tests and their relationship with reaction times: A comment on Ruchalla et al. *Intelligence* **1986**, *10*, 93–100. [[CrossRef](#)]
60. Hale, S.; Jansen, J. Global processing-time coefficients characterize individual and group differences in cognitive speed. *Psychol. Sci.* **1994**, *5*, 384–389. [[CrossRef](#)]
61. Miller, J.; Ulrich, R. Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychon. Bull. Rev.* **2013**, *20*, 819–858. [[CrossRef](#)]
62. Blotenberg, I.; Schmidt-Atzert, L. On the characteristics of sustained attention test performance—The role of the preview benefit. 2019, submitted.
63. Rodriguez, M.C.; Maeda, Y. Meta-analysis of coefficient alpha. *Psychol. Methods* **2006**, *11*, 306–322. [[CrossRef](#)]
64. Taveira, A.D.; Choi, S.D. Review study of computer input devices and older users. *Int. J. Hum. Comput. Interact.* **2009**, *25*, 455–474. [[CrossRef](#)]
65. Düker, H. *Leistungsfähigkeit und Keimdrüsenhormone*; J.A. Barth: München, Germany, 1957.
66. Krumm, S.; Schmidt-Atzert, L.; Bracht, M.; Ochs, L. Coordination as a crucial component of performance on a sustained attention test. *J. Individ. Differ.* **2011**, *32*, 117–128. [[CrossRef](#)]
67. Deary, I.J.; Stough, C. Intelligence and inspection time. *Am. Psychol.* **1996**, *51*, 599–608. [[CrossRef](#)]
68. Vernon, P.A. *Speed of Information-Processing and Intelligence*; Ablex: Norwood, NJ, USA, 1987.
69. Grudnik, J.L.; Kranzler, J.H. Meta-analysis of the relationship between intelligence and inspection time. *Intelligence* **2001**, *29*, 523–535. [[CrossRef](#)]
70. Kranzler, J.H.; Jensen, A.R. The nature of psychometric g: Unitary process or a number of independent processes? *Intelligence* **1991**, *15*, 397–422. [[CrossRef](#)]
71. Muthén, L.K.; Muthén, B.O. *Mplus User's Guide*; Muthén & Muthén: Los Angeles, CA, USA, 2015.
72. Townsend, J.T. Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychol. Sci.* **1990**, *1*, 46–54. [[CrossRef](#)]
73. Schubert, A.-L.; Hagemann, D.; Voss, A.; Schankin, A.; Bergmann, K. Decomposing the relationship between mental speed and mental abilities. *Intelligence* **2015**, *51*, 28–46. [[CrossRef](#)]
74. Stafford, T.; Gurney, K.N. Additive factors do not imply discrete processing stages: A worked example using models of the Stroop task. *Front. Psychol.* **2011**, *2*. [[CrossRef](#)] [[PubMed](#)]
75. Neubauer, A.C.; Knorr, E. Elementary cognitive processes in choice reaction time tasks and their correlations with intelligence. *Pers. Individ. Differ.* **1997**, *23*, 715–728. [[CrossRef](#)]



### **Beitrag 2:**

Blotenberg, I., & Schmidt-Atzert, L. (2019a). *On the characteristics of sustained attention test performance - The role of the preview benefit*. Manuskript eingereicht.

# **On the Characteristics of Sustained Attention Test Performance - The Role of the Preview Benefit**

Iris Blotenberg & Lothar Schmidt-Atzert

## **Authors' note**

Iris Blotenberg and Lothar Schmidt-Atzert

Philipps University, Gutenbergstr. 18, 35032 Marburg, Germany.

Correspondence concerning this article should be addressed to Iris Blotenberg

E-mail: [iris.blotenberg@staff.uni-marburg.de](mailto:iris.blotenberg@staff.uni-marburg.de).

## **Disclosure statement**

No potential conflict of interest was reported by the authors.

## Abstract

Most psychometric tests assessing sustained attention are characterized by a specific presentation mode: Many items are presented simultaneously and the test-takers are required to constantly process and react to them until the testing time is up. The aim of the present study was to look into two mechanisms that potentially underly performance in these tests: The ability to *focus* on the currently relevant item and the ability to *preprocess* upcoming items to prepare for upcoming actions. In order to assess both abilities, the d2-R test of sustained attention was modified and its stimulus arrangement (single, blocks vs. rows of stimuli) was manipulated. The measure of *focusing* was unreliable and unrelated to performance in standard sustained attention tests. However, the data indicated a strong *preview benefit*. That is, the test-takers *preprocessed* upcoming items when they got a valid preview of them, which considerably enhanced performance. Moreover, interindividual differences in the preview benefit proved to be internally reliable as well as reliable in retest and were substantially related to performance in three conventional sustained attention tests. We conclude that *preprocessing* constitutes an important component of performance in sustained attention tests and most likely represents a stable cognitive ability rather than a strategy.

Keywords: sustained attention tests, concentration tests, experimental test validation

## Introduction

Sustained attention (also referred to as concentration) may be defined as the “ability to allocate processing resources for quite a time [...] to a specific task demand” (Schweizer, 2005, p. 46). It constitutes an important precondition for many tasks in everyday and professional life (Westhoff & Hagemeister, 2005). Some of these tasks, like following a conversation or a talk, primarily require the *focusing* on the relevant stimulus (e.g. the person holding the talk) for a prolonged period of time. Others, like proofreading a mail, looking for something in the supermarket or even driving a car, also strongly benefit from a *proactive scanning of potentially relevant information* in the immediate or more distant surroundings (e.g. not only watching the street in front of you while driving, but also noticing the kids playing on the sidewalk).

Regarding psychological assessment, sustained attention is typically measured by tests in which many items are presented simultaneously and participants are required to constantly process and respond to them until the testing time is up (Krumm, Schmidt-Atzert, Schmidt, Zenses, & Stenzel, 2012; Moosbrugger & Goldhammer, 2006; Schmidt-Atzert, Krumm, & Bühner, 2008; Westhoff & Hagemeister, 2005). We postulate that this characteristic presentation mode imposes certain cognitive demands on the test-taker that correspond to the aforementioned everyday sustained attention tasks: Among them, on the one hand, the demand to *focus* on the currently relevant item, which is embedded in many simultaneously presented items (see Eriksen, 1995; Friedman & Miyake, 2004), and on the other hand, the demand to proactively *preprocess* upcoming items in order to prepare for the next actions (see Rayner, 1998; Schotter, Angele, & Rayner, 2012). In the present study, we investigate the role of *focusing* and *preprocessing* for performance in order to gain a deeper understanding of the demands and mechanisms underlying sustained attention tests.

## **Characteristics of Sustained Attention Tests**

Psychometric tests measuring sustained attention have a long tradition and have been widely used for decades (e.g. Van Der Ven, Smit, & Jansen, 1989). Typically, they require the test-takers to continuously perform a simple mental operation like searching for specific targets or check arithmetic problems. The items are presented simultaneously and participants are required to continually respond to them until the testing time is up (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm et al., 2012; Westhoff & Hagemeister, 2005). In contrast, many other attention tests (e.g. following the multidimensional attention approach; see Posner & Bois, 1971) demand the processing of single stimuli, where one stimulus is presented after another with a fixed time interval between the presentation of the next stimulus (force-paced mode).

In their studies, Krumm, Schmidt-Atzert, and Eschert (2008) and Krumm et al. (2012) demonstrated the crucial role of the presentation mode for the assessment of attention by systematically varying stimulus arrangement (single vs. many stimuli) and pace (self-paced vs. force-paced) of tests of both attention and sustained attention. They showed, for example, that different attention tests and tests of sustained attention, which usually correlate weakly to moderately, suddenly correlated highly if their presentation mode matched, indicating that the presentation mode impacts the validity of the task (Krumm, Schmidt-Atzert, & Eschert, 2008). Importantly, the self-paced processing of many items, which is typical for sustained attention tests, is considered to impose the highest cognitive demands on the test-taker as it requires continuous mental effort (Krumm et al., 2012; Schweizer & Moosbrugger, 2004; Schweizer, Moosbrugger, & Goldhammer, 2005). In the following, we more closely investigate these specific cognitive demands.

## **Demands and Mechanisms Underlying Sustained Attention Tests**

In the present study, we explore the role of two potentially relevant mechanisms for performance in sustained attention tests: Firstly, the simultaneous presentation of many items should require the *focusing* on the currently relevant item (e.g. Eriksen, 1995; Friedman & Miyake, 2004). This demand is also repeatedly mentioned in the literature as an important feature of attention in general and of sustained attention in particular (e.g. Goldhammer & Moosbrugger, 2006; Schmidt-Atzert et al., 2008). Secondly, the presentation of many items may also provide an opportunity to *preprocess* upcoming items. Indeed, a preview benefit has already been reported for an important everyday activity that also involves sustained attention – that is, reading. Eye-tracking research on reading shows that, when readers have a valid preview of the upcoming word, the preprocessing of word  $n + 1$  during a fixation of word  $n$  facilitates later processing of word  $n + 1$  (Rayner, 1998; Schotter et al., 2012). Since the ability to efficiently *preprocess* upcoming items has the potential to increase performance substantially, we presume that it could very likely be an important mechanism in sustained attention tests.

Additionally, another characteristic of concentrated attention and accordingly, another typical requirement of sustained attention tests is the demand to constantly stay on task. That is, they typically do not include short breaks like many other attention tests (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm et al., 2012). In their experimental validation of a process model of sustained attention tests, Blotenberg and Schmidt-Atzert (2019) already demonstrated, using the present data, that individual differences in the ability to handle this demand could not be reliably measured and did not predict test scores. Thus, this variable was only included here to explore potential interactions with the stimulus arrangement and the pace of the task.



## Aims of the Present Study

In order to investigate whether and to what extent *focusing* and *preprocessing* play a role in typical sustained attention tests, a modified version of the well-established d2-R test of sustained attention (electronic version, Schmidt-Atzert & Brickenkamp, 2017) was created. In the original version of this test, rows of the letters “d” and “p” with one to four marks are presented simultaneously on a computer screen and the test-taker is required to mark as many d’s with two marks (d2) as possible within a given time limit. In the modified version, this stimulus arrangement was systematically varied to impose lower (presentation of single stimuli) versus increased *focusing* demands (presentation of blocks of stimuli) and to not allow (presentation of blocks of stimuli; i.e. only one of several stimuli was relevant) versus to allow *preprocessing* (presentation of rows of stimuli, i.e. several stimuli which became relevant successively).

In order to reflect higher *focusing* demands, we would expect performance to be impeded in the conditions with blocks compared to single stimuli while a *preprocessing* mechanism should result in a substantially enhanced performance in the condition with rows (which did allow preprocessing) compared to blocks of stimuli (which did not allow preprocessing). Moreover, the pace of the task was varied to not allow (self-paced mode) versus to allow short breaks between items (force-paced mode). Finally, three standard sustained attention tests were applied to examine the relationship of the *focusing* and the *preprocessing* indices with test performance. If the two mechanisms constitute stable cognitive abilities that contribute to performance in sustained attention tests, they should be substantially correlated with performance in these tests.

## Materials and Methods

### Participants

One hundred undergraduates of the Philipp-University of Marburg participated in the study (age:  $M = 22.9$ ,  $SD = 4.6$ ,  $range = 18 - 40$ , 72 % female, 42 % studied psychology) and received partial course credit in exchange. They gave informed consent prior to participation.

### Measures

#### *Sustained Attention Tests*

A figural, numerical and verbal sustained attention test was applied to relate the *focusing* and *preprocessing* indices to performance in sustained attention tests (see Schmidt-Atzert et al., 2006). For each of these tests, the independent variable was the number of correctly marked items.

Test d2-R electronic version (Schmidt-Atzert & Brickenkamp, 2017)

In the d2-R, participants are instructed to select the letter “d” with a total of two marks out of rows of “d’s” and “p’s” with one to four marks. The task consists of 14 consecutive pages with six rows of ten stimuli per page (20 seconds per page, 4.40 minutes in total).

Revision Test (Marschner, 1980)

This paper-pencil test requires participants to assess whether a simple equation is correct and if it is, to tick it, and if it is not, to struck it out. The test consists of 15 rows with 44 items per row (30 seconds per row, 7.30 minutes in total).

Three subtests of the Berlin Intelligence Structure (BIS) test assessing verbal processing speed (Jäger, Süß, & Beauducel, 1997)

Firstly, “classification of words” requires the test-takers to strike out as many plants as possible (list of 100 words, time limit: 30 seconds). Secondly, in the task “uncompleted

words”, participants are instructed to fill in the missing letter (list of 57 words, time limit: 50 seconds). Thirdly, the task “part-whole” demands the test-takers to find as many word pairs as possible with a certain semantic relation (list of 60 words, time limit: 40 seconds).

### ***The Modified d2-R***

The modified version of the d2-R test of sustained attention was programmed in E-Prime 2.0 (Schneider, W., Eschman, A., & Zuccolotto, 2002). In this test, the letter d or p with a total of one to four marks was presented. The task was to decide whether the letter in question was a d2 and if it was, to press the button “m” (marked with a green label), with the index finger of the right hand on a German QWERTZ computer keyboard. For nontargets, the letter “c” (marked with a red label) should be pressed with the index finger of the left hand.

In the course of the six blocks of the modified d2-R, the presentation mode was systematically varied while the task remained the same. Each block consisted of 80 trials with 50 % targets (d2) and 50 % non-targets plus ten practice trials. Target and non-target switches were counterbalanced and the same stimulus never appeared twice in a row. Moreover, congruency of the flanking stimuli was counterbalanced in Blocks three to six.

In the first experimental block, single letters were presented one at a time with a 500 ms Response Stimulus Interval (RSI) between successive trials (low *focusing* demand, *preprocessing* impossible, force-paced mode). The second block was identical, except that the RSI between successive trials was zero (low *focusing* demand, *preprocessing* impossible, self-paced mode). In the third block, an additional difficulty was introduced as blocks of stimuli were presented. That is, three letters appeared next to each other and participants were required to decide whether the letter in the center was a d2 or not (this letter was marked by an arrow), while the flanking stimuli were irrelevant. In this block, the RSI between trials was 500 ms (higher *focusing* demand, *preprocessing* impossible, force-paced mode). Accordingly, in Block four, three letters appeared at a time but the RSI between successive trials was

removed (higher *focusing* demand, *preprocessing* impossible, self-paced mode). In Block five, a row of ten letters was presented and the letters became relevant one after another. The relevant letter was indicated by an arrow and the participants' task was to decide whether it was a d2 or not. After responding to that letter, there was a blank screen (RSI) for 500 ms, before the arrow moved on to the next letter in the row (higher *focusing* demand, *preprocessing* possible, force-paced mode). In the final block, a row of stimuli was presented but the RSI was removed (higher *focusing* demand, *preprocessing* possible, self-paced mode; see Figure 1 for an illustration of the procedure). Altogether, the task took about 15 min.

--- Insert Figure 1 about here ---

## Procedure

Participants were tested in groups of two to five in a laboratory. Each test session took about two hours including two ten minute-breaks. The experiment started with a short questionnaire and cognitive tasks irrelevant to the present study. After the first break, the modified d2-R, the Revision-Test and other tests not included in the present analysis were administered. Another ten minute-break was followed by the electronic version of the d2-R, the verbal sustained attention tests of the Berlin Intelligence Structure Test (BIS; classification of words, uncompleted words, part-whole) and a second exposure of the modified d2-R.

## Data Preprocessing

For the correct trials of the modified d2-R,  $z$ -values were calculated for each participant and each block and trials above  $z = 2.5$  and below  $z = -2.5$  were excluded as outliers. Moreover, accuracy below 70 % in the modified d2-R led to an exclusion of the

participant's data in this task (three participants) and accuracy below 70 % in one of the conditions led to an exclusion of the respective condition (three participants for the self-paced condition with rows of stimuli).

### ***Focusing Index***

In order to calculate the individual *focusing* index, the difference between the log-transformed<sup>1</sup> reaction times (RT) of the modified d2-R version with blocks of stimuli and the version with single stimuli was computed.

### ***Preprocessing Index***

To determine the individual differences in the extent to which the test-takers utilized the preview benefit, the difference between the log-transformed mean latencies of the modified d2-R version with rows of stimuli and the version with blocks of stimuli was calculated. Both versions demanded a *focusing* of the relevant stimulus but only the version with rows of stimuli allowed *preprocessing*.

---

<sup>1</sup> Log-transformed differences between times equal ratios. Hale and Jansen (1994) demonstrated proportionate differences in reaction times between slow and fast individuals over different tasks. We were interested in the individual's relative performance decline or improvement as a function of the presentation mode.

## Results

### Effects of Stimulus Arrangement and Pace on Performance in the Modified d2-R

To investigate the effects of stimulus arrangement and pace on RT and error rates in the modified d2-R, a repeated-measures ANOVA with the within-subjects variables *stimulus arrangement* (single, block vs. row) and *pace* (self-paced vs. force-paced) was conducted (see Figure 2). There was a main effect of *stimulus arrangement*,  $F(1.505, 140.011) = 428.883$ ,  $p < .001$ ,  $\eta^2_p = .822^2$ . Post hoc tests using the Bonferroni correction revealed that RT to rows of stimuli ( $M = 504$  ms) were significantly shorter than RT to single ( $M = 667$  ms) or blocks of stimuli ( $M = 664$  ms),  $p < .001$ , revealing a *preview benefit*. There was no significant difference between RT to single and blocks of stimuli,  $p > .95$ . Thus, surprisingly, the results did not reflect an increased *focusing demand* as there was no drop in performance in the conditions with blocks of stimuli compared to the conditions with single stimuli.

Moreover, there was a main effect of *pace*,  $F(1, 93) = 1100.875$ ,  $p < .001$ ,  $\eta^2_p = .922$ , confirming that RT in the self-paced condition ( $M = 687$  ms) were significantly longer than RT in the force-paced condition ( $M = 535$  ms). Finally, there was a significant interaction of stimulus arrangement and pace,  $F(1.366, 127.070) = 477.430$ ,  $p < .001$ ,  $\eta^2_p = .837^2$ , revealing that the difference between RT in the self-paced and the force-paced mode was especially large for rows of stimuli ( $M = 303$  ms) and smaller for single ( $M = 70$  ms) or blocks of stimuli ( $M = 83$  ms).

The error rates showed a similar pattern but were generally low, ranging from 2.7 to 4.8 % between the different versions of the modified d2-R. Descriptively, error rates were slightly lower for rows of stimuli (3.8 %) than for single (4.3 %) or blocks of stimuli (4.1 %),

---

<sup>2</sup> Mauchly's sphericity test indicated a violation of the sphericity assumption for the variable stimulus arrangement ( $\chi^2(2) = 36.633$ ,  $p < .001$ ) and for the interaction of stimulus arrangement and pace ( $\chi^2(2) = 57.333$ ,  $p < .001$ ). Therefore, the Greenhouse-Geisser correction was used to adjust degrees of freedom for the main effect of stimulus arrangement ( $\epsilon = 0.753$ ) and for the interaction of stimulus arrangement and pace ( $\epsilon = 0.683$ ).

but the main effect of *stimulus arrangement* was not significant,  $F(2, 186) = 2.488, p = .086, \eta^2_p = .026$ . Moreover, the main effect of *pace* was insignificant,  $F(1, 93) = 0.402, p = .528, \eta^2_p = .004$ . Additionally, there was a significant interaction of stimulus arrangement and pace,  $F(2, 186) = 30.900, p < .001, \eta^2_p = .249$ , indicating that for the presentation of rows of stimuli, responses in the force-paced condition were considerably less error-prone (2.7 %) than responses in the self-paced condition (4.8 %). In contrast, for the presentation of single and blocks of stimuli, the pattern was reversed and error rates were slightly higher in the force-paced condition (single: 4.6 %, block: 4.8 %) than in the self-paced condition (single: 4.1 %, block: 3.5 %). Altogether, responses were fastest and least error-prone in the force-paced condition presenting rows of stimuli.

--- Insert Figure 2 about here ---

### Descriptive Statistics and Reliability Estimates of the Study Measures

Table 1 provides the means, standard deviations and reliability estimates of the *focusing* and *preprocessing* indices and the applied sustained attention tests. The internal consistency of the indices was acceptable to good ( $r_{tt} = .78$  [*focusing*] -  $.86$  [*preprocessing*]). The retest reliability of the *preprocessing* index was good ( $r_{tt} = .81$ ), whereas the retest reliability of the *focusing* index was insufficiently low ( $r_{tt} = .24$ ). The internal consistency of the applied sustained attention tests was good to excellent ( $r_{tt} = .83$  [part-whole from the BIS] -  $.98$  [d2-R electronic version]).

--- Insert Table 1 about here ---

### Correlations of Focusing and Preprocessing with Tests of Sustained Attention

The correlations of the *focusing* and *preprocessing* indices with the applied sustained attention tests are presented in Table 2. First, it should be noted that the three sustained attention tests correlated moderately to highly with one another ( $r = .44, .57, .66$ ), despite the fact that these tasks are quite different (figural, numerical, and verbal tasks). There was no significant relationship between the *focusing* index and scores in the applied sustained attention tests. However, importantly, the *preprocessing index* correlated moderately with performance in these tests. It is noteworthy that the preprocessing index, which was determined using the modified d2-R, did not only correlate with performance in the d2-R (electronic version,  $r = .51$ ) but also with two other paper-pencil sustained attention tests ( $r = .26$  und  $.35$ , respectively).

--- Insert Table 2 about here ---



## Discussion

The present study set out to explore the role of *focusing* and *preprocessing* in typical sustained attention tests. This group of tests is characterized by the simultaneous presentation of many similar stimuli, requiring the test-takers to continuously react to them until the testing time is up (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm et al., 2012; Westhoff & Hagemester, 2005). On the one hand, given the large number of simultaneously presented items, these tests should require a rapid *focusing* on the currently relevant item among the many items. On the other hand, the presentation of many items might also provide a preview benefit as it enables the test-takers to *preprocess* upcoming items and prepare for upcoming actions. A modified version of the d2-R test of sustained attention was created and its stimulus arrangement was varied to more closely investigate both mechanisms.

First and foremost, there was a large preview benefit in the conditions that allowed *preprocessing* compared to those that did not, indicating that the participants *preprocessed* upcoming stimuli which considerably enhanced performance. Moreover, interindividual differences in the extent to which the test-takers utilized the preview benefit proved to be internally consistent, reliable in retest and to correlate substantially with performance in different sustained attention tests. Thus, an important mechanism underlying performance in sustained attention tests was revealed. Regarding *focusing*, however, our measure was unreliable and unrelated to test performance.

### The Preview Benefit in Sustained Attention Tests

Most importantly, a distinct preview benefit was observed for the presentation of rows of stimuli (which allowed a *preprocessing* of upcoming stimuli) compared to single and blocks of stimuli (which did not allow *preprocessing*), suggesting that the participants *preprocessed* upcoming stimuli which substantially facilitated performance. This preview benefit was especially large for the condition which allowed short breaks between the

presentation of stimuli (force-paced condition), indicating that these breaks reinforced *preprocessing* and substantially decreased RT and errors, further underpinning the role of *preprocessing* in this task. Moreover, a strong individual preview benefit corresponded with high performance in different sustained attention tests. This was not only true for the electronic version of the d2-R test of sustained attention but also for the numerical and verbal sustained attention tests that were presented in a paper-pencil format. Thus, in the present study, we were able to identify a novel mechanism underlying performance in sustained attention tests.

So far, we do not know much about the nature of this mechanism and its relationship with other cognitive abilities. As reported earlier, the “preview benefit” is a well-established phenomenon in research on reading and has been demonstrated in several languages (Radach & Kennedy, 2013). Moreover, Kennison and Clifton (1995) investigated a potential link between individual differences in working memory capacity and the size of the preview benefit but did not find an effect. Concerning the question whether *preprocessing* constitutes a strategy or an ability, our data do not allow a comprehensive answer. However, the high reliability of the preprocessing component suggests a stable ability rather than a strategy. For a strategy-based component of performance, one would expect a lower reliability (especially in retest), because some participants who did not utilize this strategy at the beginning should acquire and utilize this strategy in the course of the first administration or in the repeated administration, which would strongly reduce reliability.

### **The Focusing Demand in Sustained Attention Tests**

Another frequently mentioned characteristic of sustained attention in both, everyday tasks as well as in psychometric tests, is that it requires the *focusing* on the task and the relevant stimuli (e.g. Goldhammer & Moosbrugger, 2006; Schmidt-Atzert et al., 2008). In order to examine this demand and the ability to handle it, we created conditions of the

modified d2-R in which blocks of stimuli were presented and participants were required to focus on the central stimulus, while inhibiting flanking distractors. We expected this stimulus arrangement to impose higher *focusing* demands than the presentation of single stimuli, which did not require the inhibition of distractors. However, both conditions did not differ with regard to latencies or accuracy.

A possible explanation for this finding is that the *focusing* requirement in the conditions with blocks of stimuli may have been too easy. Firstly, the position of the target remained the same throughout the test. Secondly, it was marked by an arrow. Thirdly, the relevant stimulus in the center was flanked by only one distractor to the left and one to the right. The retest reliability of the focusing index was also very low ( $r_{tt} = .24$ ), suggesting that it did not capture a stable ability. Moreover, there was no relationship between the focusing index and performance in sustained attention tests. Thus, future studies concerned with focusing in sustained attention tests should possibly use a more challenging stimulus arrangement to induce a higher focusing demand.

### **Limitations and Strengths**

First of all, the current sample was drawn from a student population and thus, was not representative with regard to age, level of education and sex. Therefore, the results of the current study cannot be generalized to other samples. Moreover, we assessed individual differences in the ability to handle the characteristics of sustained attention tests using difference scores, which are typically lower in reliability than absolute scores (e.g. Miller & Ulrich, 2013). Indeed, the reliability of the *focusing* index was low. However, note that the internal consistency as well as the retest-reliability of the *preprocessing* component was above .8 and thus, proved to be good.

On the other hand, a major advantage of the present study was the systematic experimental variation of characteristic features of sustained attention tests, allowing the

isolation of the components of interest. Using experimental test validation, a deeper insight into the mechanisms underlying sustained attention test performance was gained (see also Bornstein, 2011; Borsboom, Mellenbergh, & Van Heerden, 2004).

## **Conclusion**

Using an experimental test validation approach, a novel component of performance in sustained attention tests was shown: The *preprocessing* of upcoming items. Importantly, this *preprocessing* component was substantially related to performance in different sustained attention tests that require different mental operations (e.g. searching for a target, mental arithmetics) and employ different test formats (computerized and paper-pencil format). Therefore, it is very likely that this finding can be generalized to other sustained attention and mental speed tests to the degree that they also require the continuous processing of many simultaneously presented items (see also Krumm, Schmidt-Atzert, Michalczyk, & Danthiir, 2008). It remains to be further investigated how this component relates to sustained attention in everyday life, in order to draw conclusions about whether it constitutes an inherent part of the construct or rather an ancillary effect of the way sustained attention tests are designed. From a psychometric perspective, the present study highlights the need to investigate whether the cognitive processes that a test invokes match the intended ones (see also the new *Standards for Educational and Psychological Testing*; AERA, APA, & NCME, 2014). Altogether, we believe that our findings contribute to a better understanding of the mechanisms involved in sustained attention tests.

## References

AERA, APA, & NCME (2014). *Standards for educational and psychological testing*.

Washington, DC: American Educational Research Association.

Blotenberg, I., & Schmidt-Atzert, L. (2019). Towards a process model of sustained attention tests. *Journal of Intelligence*, 7, 3.

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, 23, 532–544.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.

Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, 2, 101–118.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133, 101–135.

Goldhammer, F., & Moosbrugger, H. (2006). Aufmerksamkeit. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* (pp. 16–33). Heidelberg: Springer.

Hale, S., & Jansen, J. (1994). Global processing-time coefficients characterize individual and group differences in cognitive speed. *Psychological Science*, 5, 384 – 389.

Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test (BIS)*. Göttingen: Hogrefe.

Kennison, S. M., & Clifton, C. J. (1995). Determinants of parafoveal preview benefit in high

and low working memory capacity readers: Implications for eye movement control.

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 68–81.

Krumm, S., Schmidt-Atzert, L., & Eschert, S. (2008). Investigating the structure of attention:

How do test characteristics of paper-pencil sustained attention tests influence their relationship with other attention tests? *European Journal of Psychological Assessment*, 24, 108–116.

Krumm, S., Schmidt-Atzert, L., Michalczyk, K., & Danthiir, V. (2008). Speeded paper-pencil

sustained attention and mental speed tests. *Journal of Individual Differences*, 29, 205–216.

Krumm, S., Schmidt-Atzert, L., Schmidt, S., Zenses, E. M., & Stenzel, N. (2012). Attention

tests in different stimulus presentation modes. *Journal of Individual Differences*, 33, 146–159.

Marschner, G. (1980). *Revisions-Test*. Göttingen: Hogrefe.

Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling

reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin and Review*, 20, 819–858.

Moosbrugger, H., & Goldhammer, F. (2006). Aufmerksamkeits- und

Konzentrationsdiagnostik. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* (pp. 83–102). Heidelberg: Springer.

Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context.

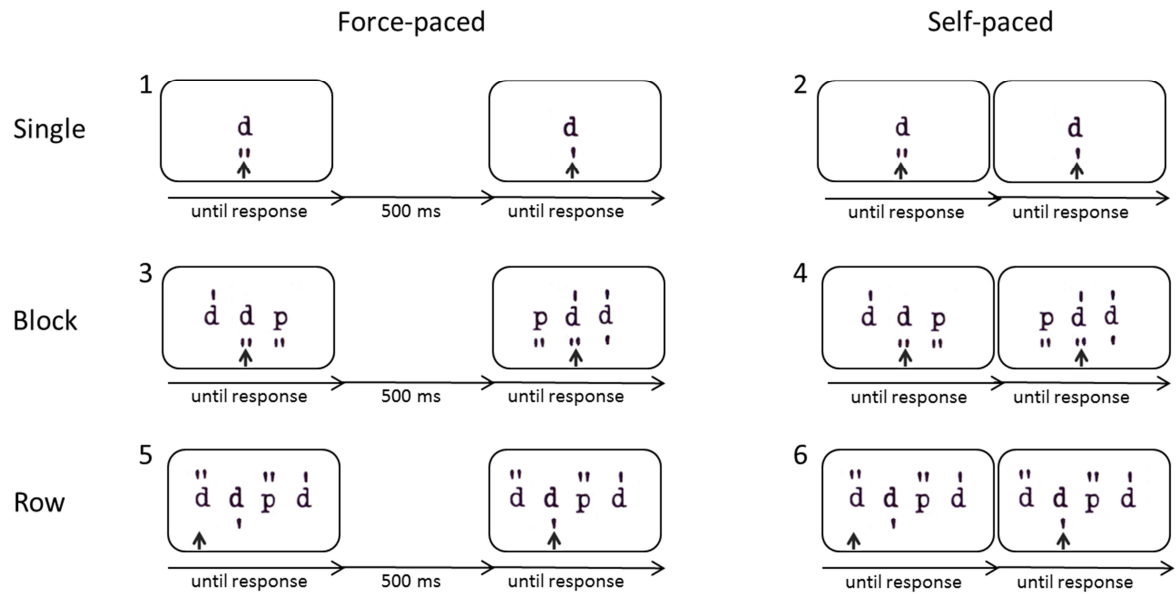
*Quarterly Journal of Experimental Psychology*, 66, 429–452.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of

research. *Psychological Bulletin*, 124, 372–422.

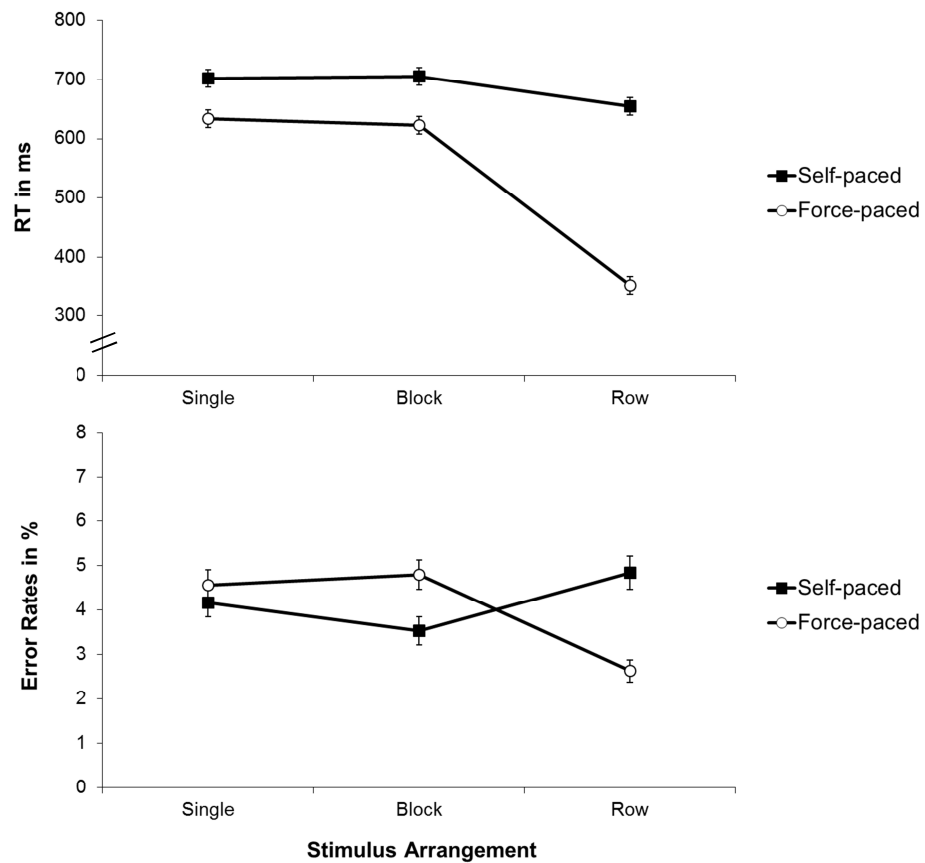
- Schmidt-Atzert, L., & Brickenkamp, R. (2017). *Test d2-R - Elektronische Fassung des Aufmerksamkeits- und Konzentrationstests d2-R*. Göttingen: Hogrefe.
- Schmidt-Atzert, L., Bühner, M., & Enders, P. (2006). Messen Konzentrationstests Konzentration? Eine Analyse der Komponenten von Konzentrationsleistungen. *Diagnostica*, 52, 33–44.
- Schmidt-Atzert, L., Krumm, S., & Bühner, M. (2008). Aufmerksamkeitsdiagnostik: Ableitung eines Strukturmodells und systematische Einordnung von Tests. *Zeitschrift für Neuropsychologie*, 19, 59–82.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime 2.0 software. Pittsburgh: Psychology Software Tools Inc.
- Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, and Psychophysics*, 74, 5–35.
- Schweizer, K. (2005). An overview of research into the cognitive basis of intelligence. *Journal of Individual Differences*, 26, 43–51.
- Schweizer, K., & Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence*, 32, 329–347.
- Schweizer, K., Moosbrugger, H., & Goldhammer, F. (2005). The structure of the relationship between attention and intelligence. *Intelligence*, 33, 589–611.
- Van Der Ven, A. H. G. S., Smit, J. C., & Jansen, R. W. T. L. (1989). Inhibition in prolonged work tasks. *Applied Psychological Measurement*, 13, 177–191.
- Westhoff, K., & Hagemeister, C. (2005). *Konzentrationsdiagnostik*. Lengerich: Pabst.

## Figures and Tables



*Figure 1.* Illustration of the six different conditions of the modified d2-R, varying stimulus arrangement (single, block, row) and pace (force-paced [with RSI] vs. self-paced [without RSI]). For all of these conditions, the task was to decide whether the relevant stimulus, indicated by an arrow, was a d2 or not and to press the corresponding key (red button for a nontarget, green button for a d2).





*Figure 2.* Mean RT (in ms) and error rates (in %) of the modified d2-R as a function of stimulus arrangement (single vs. blocks vs. rows of stimuli) and pace (self-paced vs. force-paced). Bars represent standard errors.

Table 1

*Means, standard deviations and reliability estimates of the study measures.*

Tests / Scores	<i>M</i>	<i>SD</i>	<i>r<sub>tt</sub></i>
Performance indices			
Focusing <i>RT blocks of stimuli minus RT single stimuli</i>	- 4.2 <sup>a</sup>	41.3	.24 <sup>c</sup> - .78 <sup>d</sup>
Preprocessing <i>RT rows of stimuli minus RT blocks of stimuli</i>	- 160.2 <sup>a</sup>	69.1	.81 <sup>c</sup> - .86 <sup>d</sup>
Sustained attention tests			
d2-R electronic version (figural)	220.1 <sup>b</sup>	33.9	.98 <sup>d</sup>
Revision-Test (numerical)	390.0 <sup>b</sup>	74.0	.96 <sup>d</sup>
BIS UW/CW/PW (verbal)	34.8/25.2/14.9 <sup>b</sup>	8.7/5.7/2.5	.95/.95/.83 <sup>d</sup>

*Notes.* <sup>a</sup>Difference score in ms, <sup>b</sup>Average number of correct items, <sup>c</sup>Retest reliability, <sup>d</sup>Cronbachs  $\alpha$ .

Table 2

*Correlation matrix of performance indices and sustained attention tests.*

Tests / Scores	2.	3.	4.	5.	6.
Performance Indices					
1. Focusing	-.18	-.12	-.11	-.05	.00/-.13/.00
2. Preprocessing		.50**	.35**	.26*	.12/.24*/.28**
Sustained Attention					
3. d2-R electronic version (figural)			.57**	.44**	.41**/.41**/.26**
4. Revision-Test (numerical)				.66**	.57**/.56**/.47**
5. BIS Score (verbal)					.82**/.88**/.75**
6. BIS UW/CW/PW (verbal)					

*Notes:*  $N = 94 - 100$ , due to excluded data in some tasks (pairwise deletion).

\*  $p < .05$ , \*\*  $p < .01$ .

### **Beitrag 3:**

Blotenberg, I., & Schmidt-Atzert, L. (2019b). *On the locus of the practice effect in sustained attention tests*. Manuskript eingereicht.

# **On the Locus of the Practice Effect in Sustained Attention Tests**

Iris Blotenberg & Lothar Schmidt-Atzert

## **Authors' note**

Iris Blotenberg and Lothar Schmidt-Atzert

Philipps University, Gutenbergstr. 18, 35032 Marburg, Germany.

Correspondence concerning this article should be addressed to Iris Blotenberg

E-mail: [iris.blotenberg@staff.uni-marburg.de](mailto:iris.blotenberg@staff.uni-marburg.de).

## **Disclosure statement**

No potential conflict of interest was reported by the authors.

## **Abstract**

The present study set out to explore the locus of the poorly understood but frequently reported and comparatively large practice effect in sustained attention tests. Drawing on a recently proposed process model of sustained attention tests, several cognitive tasks were administered twice in order to examine which specific component of test performance benefitted from practice and to which extent. It was shown that the tasks representing the three sub-components of sustained attention tests, namely the perception of an item, the simple mental operation to solve an item and the motor reaction to indicate a response to an item benefitted from practice. Importantly, the largest practice gain was observed for the task that required item-solving processes in addition to perceptual and motor processes. Two additional postulated mechanisms in sustained attention tests, the deliberate shifting between items and the preprocessing of upcoming items did not become more efficient through practice. Altogether, the present study suggests that the practice effect in sustained attention tests is primarily due to faster item-solving processes and to a limited extent due to a faster perception of the item as well as a faster motor response. Moreover, besides the sub-components, it is likely that also the coordination of perceptual, item-solving and motor processes benefitted from practice. Altogether, the present paper may have taken a first step towards a better understanding of the specific processes that cause the large practice gains in sustained attention tests.

Keywords: sustained attention, concentration, practice effects, process model, experimental test validation

## **Introduction**

The ability to sustain mental focus over extended periods of time, namely sustained attention (or concentration) is crucial for various everyday tasks. Accordingly, the assessment of sustained attention is relevant in many different psychological domains like neuro-, clinical or traffic psychology as well as in personnel selection (Westhoff & Hagemeister, 2005). Sustained attention tests' excellent psychometric properties have been demonstrated in multiple studies (e.g. Brickenkamp, Schmidt-Atzert, & Liepmann, 2010; Mirsky, Anthony, Duncan, Ahearn, & Kellam, 1991; Schmidt-Atzert & Bühner, 2000; Schmidt-Atzert, Bühner, & Enders, 2006). However, there are, like in other cognitive ability tests, substantial practice effects in this group of tests, for which the causes are largely unknown (Bühner, Ziegler, Bohnes, & Lauterbach, 2006; Büttner & Schmidt-Atzert, 2004). The aim of the present study was to explore the locus of the practice effect in these tests by drawing on a recently proposed process model of sustained attention tests.

## **Practice Effects in Sustained Attention Tests**

Practice or retest effects, that is, a test score gain after prior exposure to the same or an alternate form of a test under comparable conditions (Lievens, Reeve, & Heggstad, 2007), is a well-documented phenomenon for cognitive ability tests (Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007; Randall & Villado, 2017; Scharfen, Peters, & Holling, 2018). In a recent meta-analysis, Scharfen, Peters, et al. (2018) revealed practice effects of a third standard deviation for cognitive ability tests in general and a slightly higher practice effect of .37 standard deviations for tests of processing speed, which are conceptually similar, and empirically indistinguishable from sustained attention tests (Krumm, Schmidt-Atzert, Michalczyk, & Danthiir, 2008; Schmidt-Atzert et al., 2006). Other studies not considered in the meta-analysis reported even larger practice effects between half and one standard deviation from the first to the second test administration (Bühner et al., 2006; Westhoff &

Dewald, 1990). For the German sustained attention test d2, which was applied in the present study, practice effects ranged from two-thirds standard deviation for the paper-pencil version and four-fifths standard deviation for the latest electronic version (Bühner et al., 2006; Schmidt-Atzert & Brickenkamp, 2017; Steinborn, Langner, Flehmig, & Huestegge, 2017). Moreover, several studies showed increasing practice effects over several test administrations. For example, after eleven test administrations, Westhoff and Dewald (1990) revealed large practice effects of nearly three standard deviations for a figural and around two and a half standard deviations for a numerical sustained attention test. Even after these many repetitions, though test scores increased more slowly, they had not reached a plateau yet (see also Scharfen, Blum, & Holling, 2018). Additionally, while retest effects have been shown to decline with longer retest intervals, they have also been reported to decline rather slowly, so that it takes five years for them to vanish (Scharfen, Peters, et al., 2018). Thus, practice effects in cognitive ability tests in general and in sustained attention tests in particular deserve consideration. Importantly, retesting is quite common in many contexts like personnel selection or the educational sector, but also in neuropsychology (Scharfen, Peters, et al., 2018) and it impacts the validity of the task (Lievens et al., 2007; Scharfen, Peters, et al., 2018; Te Nijenhuis, van Vianen, & van der Flier, 2007).

Although various factors are being discussed (Lievens, Buyse, & Sackett, 2005; Lievens et al., 2007; Randall & Villado, 2017), the causes and locus (that is, the specific processes that become more efficient through practice) of the practice effect in sustained attention tests are not comprehended in detail (Büttner & Schmidt-Atzert, 2004; Scharfen, Blum, et al., 2018). In fact, there is a paucity of studies that investigate how retesting affects the specific processes and mechanisms involved in these tests. In the present study, we address this knowledge gap by focusing on the components that have been shown to drive performance in sustained attention tests. That is, using an approach of experimental test validation, we examine which of these components benefit from practice and to which extent.



In the following, the characteristics of sustained attention tests and a recently proposed process model are introduced.

### **A Process Model of Sustained Attention Tests**

For more than a hundred years now, sustained attention is measured using simple, homogenous stimuli and comparatively easy tasks, like letter cancellation, simple mental arithmetics or sorting according to categories (Westhoff & Graubner, 2003; Westhoff & Hagemeister, 2005). Importantly, even more critical than the task itself is the typical presentation mode of these tests, that is, many stimuli are presented at the same time and the participants are required to continuously work and respond to them until the test is over (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm, Schmidt-Atzert, Schmidt, Zenses, & Stenzel, 2012; Westhoff & Hagemeister, 2005). As a main indicator of performance, the number of marked items (speed) or the number of correctly marked items (error-corrected speed) is assessed (Schmidt-Atzert et al., 2006).

Blotenberg and Schmidt-Atzert (2019b) recently proposed a process model of sustained attention tests. They argued that the operations required in these tests on the item-level are comparatively straightforward. Independent of whether the task requires the cancellation of targets, mental arithmetics or the sorting according to categories, they all require a fast *perception* of the item, which is followed by a *simple mental operation* to solve the item and a *motor response*. Additionally, according to Blotenberg and Schmidt-Atzert (2019a) the model is to be extended when considering the characteristic presentation mode of sustained attention tests, that demands the deliberate, self-paced shifting between many simultaneously presented items (Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm et al., 2012; Westhoff & Hagemeister, 2005). This characteristic presentation mode should require the test-taker to shift from one item to another as quickly as possible (*item shifting*, Krumm, Schmidt-Atzert, & Eschert, 2008; Krumm et al., 2012; Schweizer & Moosbrugger, 2004; Van

Breukelen et al., 1995; Westhoff & Hagemester, 2005). Moreover, it should also demand the test-taker to *focus* on the currently relevant item while ignoring the surrounding items (see also Eriksen, 1995; Friedman & Miyake, 2004). Finally, the simultaneous presentation of many items might also provide the opportunity to *preprocess* upcoming items in order to prepare for upcoming actions (see also Rayner, 1998; Schotter, Angele, & Rayner, 2012).

In two studies, that were partly based on the same data as the current study, it was demonstrated that *perceptual*, *mental operation* and *motor speed* already explained a large amount of variance in conventional sustained attention tests, namely 55 - 74 % (Blotenberg & Schmidt-Atzert, 2019b). Importantly, *perceptual* and *mental operation speed* were the strongest predictors of test performance, while there was a consistent trend towards a minor influence of *motor speed*. Moreover, while there were considerable *item shifting costs* for the deliberate, self-paced shifting between items compared to a presentation mode which involved short intervals between successive items (also called the force-paced mode), these were not related to performance in sustained attention tests. Additionally, there was no effect of *focusing*, that is, performance did not significantly differ between conditions which should have induced higher versus lower focusing demands (Blotenberg & Schmidt-Atzert, 2019a). Thus, it seems that the manipulation of an increased focusing demand in the modified d2 was not successful and therefore, it will not be further examined in the current study. Finally, the authors found a large preview benefit, that is, performance was considerably facilitated when the test-takers received the opportunity to preprocess upcoming stimuli. This *preprocessing* component proved to be substantially correlated with performance in sustained attention tests (Blotenberg & Schmidt-Atzert, 2019a).

## **The Present Study**

Drawing on these findings, the aim of the present study was to more closely investigate the practice effect in sustained attention tests and more specifically, to examine

which sub-components benefit from the repeated test administration. First of all, a modified version of the d2-R test of sustained attention was created, which closely resembled the original d2-R but allowed for the measurement of reactions to single items. This so-called “model d2” was administered twice and we expected practice gains of a similar magnitude as in the original d2, that is, of around two-thirds a standard deviation (Bühner et al., 2006). To assess the practice effects of *perceptual*, *mental operation* and *motor speed*, several cognitive tasks were selected so that they would successively demand additional sub-components: The inspection time task was administered as a prototypical measure of *perceptual* speed (Brand & Deary, 1982; Schweizer & Koch, 2003; Vickers, Nettelbeck, & Willson, 1972). The simple reaction time task, which is considered to impose mainly motor (Ackerman, 1988; Fleishman & Hempel, 1954; Sternberg, Monsell, Knoll, & Wright, 1978) but also basic perceptual demands (Neubauer & Knorr, 1997), was applied to assess *motor speed*. Additionally, a simple version of the modified d2, in which only single stimuli were presented one after another including short breaks, was applied to measure *mental operation speed* beyond perceptual and motor processes. All of these tasks were applied twice and we expected substantial practice gains for at least one of the three tasks and the respective sub-components. We then investigated the practice effects of *perceptual*, *mental operation* and *motor speed* and their relationship with the practice effect in the model d2.

Furthermore, we explored how retesting affected the *item shifting* and *preprocessing* mechanisms. In other words: Is the large practice effect in sustained attention tests (partly) due to a more efficient *shifting between items* or *preprocessing of upcoming items* in the repeated test administration? To address this research question, further modified versions of the d2 were created and administered twice. It was examined how practice affected performance in the different versions with varying pace (self-paced vs. force-paced) and stimulus arrangements (single, blocks vs. rows of stimuli). For an effect of practice on *item shifting* we should find an interaction of practice and pace so that the difference between the

self-paced and the force-paced conditions shrunk through practice. Such an interaction would suggest that practice reinforced the participants' capability to handle the intentional *item shifting* required when they work through the test in a self-paced manner. Additionally, if practice reinforced *preprocessing*, we should observe a significantly larger practice effect for the conditions with rows of stimuli (which allowed preprocessing) compared to the conditions with single or blocks of stimuli (which did not allow preprocessing).

## **Method**

### **Participants**

One hundred undergraduates (72 female, 42 studied psychology) with a mean age of 22.9 years ( $SD = 4.6$ ,  $range = 18 - 40$ ), participated in the present study and received partial course credit in exchange. They gave informed consent prior to participation in accordance with the Declaration of Helsinki.

### **Measures**

#### Perceptual Speed - The Inspection Time Task

In the inspection time task, a pi-shaped figure with two legs of markedly different lengths is presented and the participant is required to perceive and indicate which of the two lines is longer (Vickers & Smith, 1986). Based on the response accuracy, the figure's exposure time is adjusted adaptively in order to assess the shortest time necessary to correctly perceive it. The inspection time task is considered a primarily perceptual task (Brand & Deary, 1982; Schweizer & Koch, 2003; Vickers et al., 1972) and performance in this task was utilized as a measure of the sub-component of perceptual speed. The present version of the task was programmed in E-Prime 2.0 (Schneider, W., Eschman, A., & Zuccolotto, 2002). Each trial started with a white fixation cross against a black background (1000 ms). It was

followed by a Pi-shaped figure with one shorter (3.5 cm) and one longer leg (4.5 cm). Afterwards, the legs were covered by a backward mask to prevent processing from stored traces (300 ms). The subject's task was to indicate which leg was longer by pressing "c" for the left leg (with the left index finger) and "m" for the right leg (with the right index finger) on a German QWERTZ computer keyboard. The exposure time of the figure was varied using a staircase procedure, i.e. four correct responses led to a shortened exposure time and an incorrect response led to an increase in exposure time. After three practice trials, the staircase procedure started with an exposure time of 157 ms and decreased in steps of 66 ms (at the beginning) to 16.5 ms (in the course of the experiment). The experiment ended after 15 reversals (i.e. when the exposure time suddenly increases after it had been decreasing before or vice versa) or a maximum of 96 trials. The individual inspection time was the shortest exposure time to which the participant answered to correctly for four times in a row.

#### Motor Speed - The Simple Reaction Time Task

As a prototypical measure of motor speed, a simple reaction time task was applied (Ackerman, 1988; Fleishman & Hempel, 1954; Sternberg et al., 1978). Participants were instructed to press the button "c" (20 trials, left index finger) and afterwards "m" (20 trials, right index finger) as fast as possible as soon as a black dot (2 x 2 cm) appeared on screen. The dot was presented until a response was made and the next dot appeared after 1000 ms (+/- 100 ms jitter). The simple reaction time task consisted of 40 experimental plus 10 practice trials. The dependent variable was the mean reaction time.

#### The Modified d2

For the purpose of the present study, several modified versions of the d2-R test of sustained attention were created in E-Prime 2.0 (Schneider, W., Eschman, A., & Zuccolotto, 2002). All of these versions presented the letter "d" or "p" with one to four marks above and below it. The test-taker's task was to decide whether the respective letter was a target, that is,

a “d” with a total of two marks, or a nontarget and to press the key “c” for nontargets (left index finger; the key was colored red) or “m” for targets (right index finger; the key was colored green) on a German QWERTZ computer keyboard. While this task remained the same throughout the different versions, we varied the presentation mode to systematically manipulate characteristic features of sustained attention tests, i.e. the stimulus arrangement and the pace.

Altogether, there were six blocks and each block consisted of 80 stimuli plus 10 practice trials. The whole task took about 15 minutes. In the first block, “single stimuli, force-paced”, a single letter (with marks) was presented at a time and the participant had to decide whether the respective letter was a d2 or not and to press the respective key. After the test-takers’ response, there was a 500 ms response-stimulus interval before the next letter appeared. This simple modified version of the d2 was created to measure the speed of the item solving process (*mental operation speed*) beyond perceptual and motor speed. Importantly, as single stimuli were presented and short intervals were included between successive stimuli, it did not require further critical mechanisms of sustained attention tests such as item shifting or preprocessing. In the second block, “single stimuli, self-paced”, a single letter was presented at a time but the response-stimulus interval was removed, so that this condition should require the immediate *shifting between items*. In Block three, “blocks of stimuli, force-paced”, three stimuli were presented at a time, but only the one in the center was relevant. After the test-taker made a response, there was a 500 ms response-stimulus interval before the next three letters appeared. Similarly, in Block four, “blocks of stimuli, self-paced”, three stimuli were presented at a time but this time, the response-stimulus interval was removed so that participants had to immediately *shift to the next item*. In Block five, “rows of stimuli, force-paced”, rows of stimuli were presented and they became relevant one after another, allowing the *preprocessing* of upcoming stimuli. An arrow indicated the currently relevant stimulus and after responding to it, the screen went blank for a 500 ms

response-stimulus interval until the stimuli appeared again and the arrow moved on to the next stimulus. Similarly, in Block six “rows of stimuli, self-paced”, rows of stimuli were presented which became relevant successively and therefore, allowed *preprocessing*. However, this time, the response-stimulus interval was removed so that the arrow moved on to the next item right after the test-takers’ response, requiring the test-taker to immediately shift to the next item. This final condition was designed to closely resemble the original presentation mode of the d2-R test of sustained attention and is referred to as the “model d2”.

--- Insert Figure 1 about here ---

#### Test d2-R Electronic Version (Schmidt-Atzert & Brickenkamp, 2017)

The d2-R (electronic version) was applied as a conventional measure of sustained attention test performance. In the d2-R, the task is to select the letter “d” with a total of two marks out of many “d”s and “p”s with one to four marks by clicking on the respective letter with a computer mouse. 14 different screens with sixty letters per screen (structured in six rows of ten letters per row) were presented and the participants had to mark as many d2’s as possible within 20 seconds. The dependent variable was the number of correctly marked items minus the number of confusion errors (error-corrected speed). Altogether, the task took 4.40 minutes plus instructions.

#### **Procedure**

Participants were tested in groups of two to five in a laboratory. Each test session took about two hours and included two ten minute-breaks. The experiment started with a short questionnaire and cognitive tests irrelevant to the present study. After the first break, the inspection time task, the simple reaction time task, the modified d2 and further cognitive tasks were administered and the participants took another ten-minute break. Then, the electronic

version of the d2-R and further cognitive tasks were presented. Finally, participants performed the inspection time task, the simple reaction time task and the modified d2 another time and completed a short questionnaire to assess strategies and problems during the test session.

### **Data Preprocessing**

Participants who stated difficulties with the inspection time task in the post-experimental questionnaire and whose inspection time was above  $z = 4$  were excluded as outliers (three participants in the first, two in the second administration). Participants whose accuracy in the modified d2 was below 70 % were excluded from the analysis (three participants in the first, none in the second administration). Additionally, accuracy below 70 % in one of the conditions of the modified d2 led to an exclusion of the respective condition (three participants in the self-paced condition with rows of stimuli for both administrations).

For the analysis of RT in the modified d2, errors were excluded (first administration: 5.3 %, second administration: 5.3 %) and for the correct trials, z-values were calculated for each participant in each block. Trials with RT above  $z = 2.5$  and below  $z = -2.5$  were excluded as outliers (first administration: 3.1 %, second administration: 3.0 %). The same cut-off values were applied for outliers in the simple reaction time task (first administration: 6.9 %, second administration: 7.8 %).

### **Analysis Strategy**

First of all, it was examined whether RT in the model d2, that was designed to resemble the original d2, showed practice gains of the expected size reported in earlier studies. Secondly, in order to identify the locus of the practice effect within the proposed process model, we investigated which of the tasks that were applied to measure the sub-components of sustained attention performance, namely *perceptual*, *mental operation* or



*motor speed* benefitted from practice and to which extent. Therefore, we computed the practice effects in the respective tasks, that is, the practice effect in the inspection time task, the simple reaction time task and the simple modified d2. Moreover, we explored whether the practice effects in the respective sub-components were significantly related to the practice effect in the model d2.

Thirdly, we turned towards further postulated mechanisms in sustained attention test performance; *item shifting* and *preprocessing*. It was investigated to what extent they benefitted from practice. Therefore, for RT and error rates, a repeated-measures ANOVA with the within-subject factors practice (first vs. second administration) and pace (self-paced vs. force-paced), as well as practice (first vs. second administration) and stimulus arrangement (single vs. blocks vs. rows of stimuli) was conducted and it was examined whether the stimulus arrangement or pace interacted with the practice of the task. For an effect of practice on *item shifting*, there should be an interaction of practice and pace so that the difference between the self-paced and the force-paced mode became smaller through practice. With regard to *preprocessing*, there should be an interaction of practice and stimulus arrangement, indicating that the preview benefit became larger through practice.

## Results

### Descriptive Statistics and Practice Effects

Table 1 provides the means, standard deviations, retest reliabilities, the practice effects in ms and effect sizes of the practice effects for the model d2 and the tasks assessing the sub-components of the process model. Firstly, it was assessed whether RT in the model d2 reflected a substantial practice gain from the first to the second administration. This version of the modified d2 was designed to resemble the original version of the d2-R test of sustained attention. Indeed, this version of the modified d2 correlated between  $r = .68$  (first

administration of the model d2) and  $r = .73$  (second administration of the model d2) and thus, highly with performance in the original d2-R (electronic version). Importantly, it did show a large and significant practice effect,  $t(93) = 13.221$ ,  $p < .001$ . The effect size was nearly two-thirds standard deviation ( $Glass' \Delta = 0.62$ ) and therefore similar in size as the typical practice effect in standard sustained attention tests (Bühner et al., 2006; Westhoff & Dewald, 1990) and the original d2-R (electronic version,  $Glass \Delta = .81$ )

In order to investigate the effects of practice on the sub-components, a repeated-measures ANOVA with the within-subjects variables practice (first vs. second administration) and task (inspection time task, simple reaction time task vs. simple modified d2) was conducted. There was a main effect of practice,  $F(1, 91) = 234.421$ ,  $p < .001$ ,  $\eta^2_p = .720$ , demonstrating a strong practice effect in the three subtasks. Additionally, significant practice effects were confirmed for each of the three tasks by conducting separate t-tests ( $p < .01$ ). Moreover, there was a main effect of the task itself,  $F(1.339, 121.890) = 4574.412$ ,  $p < .001$ ,  $\eta^2_p = .980^1$  and an interaction of practice and the task,  $F(1.401, 127.473) = 195.263$ ,  $p < .001$ ,  $\eta^2_p = .682$ . Of the three tasks, the simple modified d2 benefitted most from practice (measure of *mental operation speed* beyond perceptual and motor processes;  $M = -97$  ms), followed by considerably smaller practice effects in the simple reaction time task (measure of *motor speed* beyond basic perceptual processes;  $M = -12$  ms) and the inspection time task (measure of *perceptual speed*;  $M = -6$  ms).

For the assessment of the sub-components, the selection of the tasks was based on the assumption that they would successively include additional sub-components of sustained attention tests (that is, the inspection time task as a measure of *perceptual speed*, the simple reaction time task as a measure of *motor speed* and basic perceptual processes, and the modified d2 as a measure of *mental operation speed* beyond perceptual and motor processes).

---

<sup>1</sup> Mauchly's sphericity test indicated a violation of the sphericity assumption for the variable task ( $\chi^2(2) = 61.158$ ,  $p < .001$ ) and for the interaction of practice and task ( $\chi^2(2) = 50.237$ ,  $p < .001$ ). Therefore, the Greenhouse-Geisser correction was used to adjust degrees of freedom for the main effect of task ( $\epsilon = 0.670$ ) and for the interaction of practice and task ( $\epsilon = 0.700$ ).

Please note that the following analyses are based on interindividual differences rather than on means. For stepwise multiple regression analyses displaying the incremental validity of these tasks for the prediction of performance in the original d2-R test of sustained attention (Schmidt-Atzert & Brickenkamp, 2017), see Appendix A, Table A1. In line with the assumption, there was an incremental increase in explained variance for each task that was applied to measure an additional sub-component: At t1 (t2), the inspection time task explained 7% (9%) of the d2-R variance. When the simple reaction time task was included in the model, it added another 7% (3%), and finally, the inclusion of the modified d2 increased the explained variance by 34% (29%).

--- Insert Table 1 about here ---

### **Correlations of the Practice Effects in the Model d2 and the Sub-Components**

Secondly, correlations between the practice effect in the model d2 and the practice effects in the tasks assessing the sub-components were examined (see Table 2). In order to determine the respective practice effects, for each task, performance in the first administration was partialled out of the performance in the second administration so that the remaining residual captured the change from the first to the second administration for each task (and measurement error). Only the practice effect in the simple modified d2 was significantly related to performance in the model d2. However, the internal consistency of these residuals was low for some of the tasks ( $r = .47$  for the model d2 but  $r = .91$  for the simple modified d2), which is not surprising, as they were derived by partialling out considerable systematic variance. When the correction for attenuation was applied to rid this correlation of its measurement error (Spearman, 1904), this correlation increased to .50.

--- Insert Table 2 about here ---

## Effects of Practice on Item Shifting and Preprocessing

### RT Analysis

In order to assess the effects of practice on *item shifting*, a repeated-measures ANOVA with the within-subject variables *practice* (first vs. second administration) and *pace* (self-paced vs. force-paced) was conducted for RT in the different conditions of the modified d2 (see Figure 2). As expected, it revealed a significant effect of *practice*,  $F(1, 93) = 457.033$ ,  $p < .001$ ,  $\eta^2_p = .831$ , confirming that RT decreased from the first ( $M = 611$  ms) to the second ( $M = 521$  ms) administration. Additionally, there was a significant effect of *pace*,  $F(1, 93) = 2186.440$ ,  $p < .001$ ,  $\eta^2_p = .959$ . That is, the self-paced mode ( $M = 642$  ms) impeded performance compared to the force-paced presentation of items ( $M = 490$  ms). Finally, there was no significant *interaction between practice and pace*,  $F(1, 93) = .000$ ,  $p = .990$ ,  $\eta^2_p = .000$ , indicating that the difference between the self-paced and the force-paced mode did not significantly change from the first ( $M_{RT\text{ Difference}} = 152$  ms) to the second administration ( $M_{RT\text{ Difference}} = 152$  ms). Thus, the results do not suggest that item shifting benefitted from practice.

A repeated-measures ANOVA with the within-subject variables *practice* (first vs. second administration) and *stimulus arrangement* (single, block vs. row) was run to investigate the effects of practice on *preprocessing*. Besides the main effect of *practice*,  $F(1, 93) = 457.033$ ,  $p < .001$ ,  $\eta^2_p = .831$ , the main effect of *stimulus arrangement* was significant,  $F(2, 186) = 643.587$ ,  $p < .001$ ,  $\eta^2_p = .874$ . Post hoc tests using the Bonferroni correction showed that RT to rows of stimuli ( $M = 457$  ms) were significantly shorter than RT to single ( $M = 614$  ms) or blocks of stimuli ( $M = 628$  ms),  $p < .001$ , revealing a strong *preview benefit*. Moreover, the interaction of practice and stimulus arrangement is crucial to assess the effects of practice on this preview benefit, since preview is restricted to the presentation of rows of

stimuli. The interaction was significant,  $F(2, 186) = 19.411, p < .001, \eta^2_p = .173$ . However, post hoc tests using the Bonferroni correction showed that the practice effect was larger for the presentation of single stimuli (*mean practice effect* = 105 ms) than for the presentation of rows of stimuli (*mean practice effect* = 93 ms), which then again was larger than the practice effect for blocks of stimuli (*mean practice effect* = 72 ms). Only the conditions presenting rows of stimuli allowed preprocessing but the practice effect in these conditions was smaller than the practice effect in the conditions with single stimuli, which did not allow preprocessing. Altogether, the data do not indicate that practice led to a substantially enlarged preview benefit.

#### Error Rates

Error rates were generally low (2.3 - 6.1 %) and the analyses yielded small effect sizes for the main effects and interactions. The repeated-measures ANOVA with the within-subjects variables *practice* and *pace* revealed a significant main effect of *practice*,  $F(1, 93) = 10.850, p = .001, \eta^2_p = .104$ . Surprisingly, error rates were slightly higher for the second administration of the modified d2 ( $M = 4.6$  %) than for the first ( $M = 4.1$  %), indicating a speed-accuracy trade-off in the second administration. The main effect of *pace* and the interaction of *practice* and *pace* were not significant,  $F(1, 93) < 1, p > .3, \eta^2_p < .02$ . Thus, the results do not indicate that *item shifting* benefitted from practice with respect to accuracy.

For the repeated-measures ANOVA with the within-subjects variables *practice* and *stimulus arrangement*, besides the main effect of *practice*,  $F(1, 93) = 10.850, p = .001, \eta^2_p = .104$ , there was also a main effect of *stimulus arrangement*,  $F(2, 186) = 16.663, p < .001, \eta^2_p = .152$ . Post hoc tests using the Bonferroni correction showed that error rates were significantly lower for the conditions with rows of stimuli ( $M = 3.7$  %) than for the conditions with single ( $M = 4.8$  %) or blocks of stimuli ( $M = 4.6$  %,  $p < .001$ ), indicating a *preview benefit* for rows of stimuli. Finally, there was a significant interaction of *practice* and *stimulus arrangement*,

$F(2, 186) = 7.448, p = .001, \eta^2_p = .074$ . This interaction showed that, for the presentation of single and blocks of stimuli, error rates increased from the first to the second administration ( $1^{st} M_{single} = 4.3\%$  and  $M_{blocks} = 4.1\%$ ,  $2^{nd} M_{single} = 5.2\%$  and  $M_{blocks} = 5.1\%$ ) while they stayed roughly the same for rows of stimuli ( $1^{st} M_{rows} = 3.8\%$ ,  $2^{nd} M_{rows} = 3.6\%$ ). Altogether, the data did not indicate that error rates decreased with practice for the presentation of rows of stimuli, suggesting that the preview benefit did not markedly benefit from practice.

--- Insert Figure 2 about here ---

## Discussion

The aim of the present study was to explore the locus of the practice effect in sustained attention tests from two different angles of approach: First, drawing on a recently proposed process model of sustained attention tests, we investigated which of the subcomponents of performance, namely *perceptual*, *mental operation* and *motor speed*, benefitted from practice and to which extent. Second, we looked into further mechanisms that have been shown to play a role in sustained attention tests, namely *item shifting* and *preprocessing*, and examined whether they became more efficient through practice. Subtasks that assessed the speed in the sub-components of sustained attention tests and modified versions of the d2 test of sustained attention were applied twice to obtain measures of the respective sub-components and the *item shifting* and *preprocessing* mechanisms in the first and second test administration.

Regarding the process model, it was shown that all the three tasks that assessed the sub-components of sustained attention tests, i.e. the *perception* of the item, a *simple mental operation* to solve the item and the *motor reaction* to it, benefitted from practice. However, while effect sizes of the practice effect were small for the measures of *perceptual* and *motor speed*, the practice effect was large for the task that additionally required item solving

processes. Thus, the two additional demands involved in this task, that is, the *simple mental operation* to solve the item and possibly also the *coordination* of the three sub-components seem to especially benefit from practice. Moreover, of the practice effects in the three sub-components, only the practice effect in this task was substantially correlated with the practice effect in the model d2. Regarding the preprocessing of upcoming items and the deliberate shifting between items, the results do not support a practice effect for these postulated mechanisms of sustained attention tests. Thus, the large practice effect in sustained attention tests cannot be attributed to a more efficient *item shifting* or *preprocessing*.

### **Effects of Practice on the Sub-Components**

The first aim of the present study was to explore which of the sub-components of sustained attention tests benefitted from practice and could therefore cause the large practice effect in this group of tests. First of all, it was confirmed that the model d2, which was created to closely resemble the original d2, correlated highly with the original d2-R ( $r = .68$  in the first to  $r = .73$  in the second administration) and showed a practice effect within the typical range of conventional sustained attention tests (two-thirds standard deviation, Bühner et al., 2006).

Regarding the tasks assessing the sub-components of the process model, all of them showed practice gains. However, effect sizes were rather small for the inspection time and simple reaction time task, which were applied to obtain measures of *perceptual* (Vickers et al., 1972) and *motor* speed (Sternberg et al., 1978). That means, getting used to the stimulus material and the input device, which leads to a more efficient *perception* and *motor reaction*, enhances the test-takers' performance only to a limited extent. However, note that our sample was young and very skilled in using a computer keyboard. Thus, with regard to motor speed, it is conceivable that the practice effect is considerably larger for elderly test-takers who are less used to a computer keyboard.

Strikingly, the practice effect was large for the simple modified d2, which involved all of the three sub-components of sustained attention tests, namely perceptual, mental operation and motor speed. Moreover, the practice effect in this task, assessed as the residual of the second administration after the first administration of the test was partialled out, was substantially correlated with the practice effect in the model d2. The size of the practice effect in this task and the correlation with the practice effect in the model d2 suggests that the additional requirement of this task, that is, the speeded processing and decision on the correct answer to the item, seems to especially benefit from practice. Hence, the relatively basic decisions required in sustained attention tests (e.g. deciding whether a stimulus is a d2 or not, whether numbers add up or not etc.) seem to become more efficient through the repeated administration and to speed up reactions substantially.

Beyond that, this task also imposed higher demands on the coordination of action patterns than the other tasks. The concept of coordination, that is, the organisation, efficient timing and execution of action patterns, has long been considered a key demand of sustained attention tests (Düker, 1957; Krumm, Schmidt-Atzert, Bracht, & Ochs, 2011; Westhoff & Hagemeyer, 2005). As the simple modified d2 required all of the three sub-components, it also imposed higher demands on their coordination than the less complex measures of perceptual and motor speed. Hence, when the responses to the simple modified d2 speeded up, this could partly be due to a more efficient coordination of the sub-components. In an earlier study, Krumm et al. (2011) had demonstrated an influence of the coordination of action patterns on the practice effect in a sustained attention tests.

### **Effects of Practice on Preprocessing and Item Shifting**

The second aim of the present study was to investigate whether two postulated mechanisms in sustained attention tests, namely *item shifting* and *preprocessing*, benefitted from practice and could therefore cause the large practice effect in these tests. In order to do



so, a modified version of the d2-R test of sustained attention was created in which the pace (self-paced vs. force-paced) was varied so that it either required the deliberate, self-paced *shifting between items* (self-paced conditions) or not (force-paced conditions). Additionally, the stimulus arrangement (single blocks vs. rows of stimuli) was varied to not allow (single or blocks of stimuli) versus to allow *preprocessing* (rows of stimuli, which became successively relevant), respectively. These different conditions of the modified d2 were administered twice in order to explore whether *item shifting* or *preprocessing* became more efficient through practice.

However, the data did not indicate that practice had a positive effect on either *item shifting* or *preprocessing*. Regarding *item shifting*, there was no significant interaction of practice and pace and the difference between the self-paced and the force-paced condition remained largely unchanged throughout both administrations. With regard to *preprocessing*, while there was an interaction of practice and stimulus arrangement, it indicated that the practice effect for single stimuli (which did not allow preprocessing) was even larger than for rows of stimuli (which allowed preprocessing), suggesting that practice did not reinforce preprocessing. Indeed, the relative size of the preview benefit effect remained the same throughout both administrations. Altogether, while the present data reflected the role of *item shifting* and *preprocessing* in sustained attention tests, they did not indicate that either of these mechanisms benefitted from practice. Quite the contrary, these results suggest that the large practice effect in sustained attention tests is most likely not due to a more efficient *item shifting* or *preprocessing* in retest.

### **Limitations and Strenghts**

It is important to acknowledge the limitations of the present study. First, the present sample was not representative with regard to age, sex or educational level and the present findings cannot be generalized to other samples. That is, our sample mostly consisted of

young and female students, who are, for example, typically very skilled in using a computer and a computer keyboard.

Another weakness of the present study was that the test-takers extensively practiced the d2 and similar tasks, that is, they completed several conditions of the modified d2 twice and additionally carried out the original test d2-R, so that the practice effects might have been overestimated. Moreover, the model d2, which was created to resemble the original d2, and the simple modified d2 overlapped quite a bit, so that the correlation of the practice effect in both tasks might have been inflated due to their similarity.

Furthermore, there was a small speed-accuracy trade-off with regard to the practice effect in the modified d2. That is, while reaction times considerably decreased from the first to the second administration, there was a slight increase in the error rates. Thus, it seems that the practice effect in the second administration of the modified d2 was partly due to a style of working that focused more on speed than on accuracy. However, note that the size of the reversed practice effect in the error rates was descriptively small (error rates of 4.1 % in the first and 4.6 % in the second administration) and also moderate in its effect size ( $\eta^2_p = .104$ ), while the practice effect in the reaction times was very large (1<sup>st</sup> administration:  $M = 611$  ms; 2<sup>nd</sup> administration:  $M = 521$  ms;  $\eta^2_p = .831$ ).

Finally, we investigated the effects of practice on the sub-components of sustained attention tests using measures of perceptual, mental operation and motor speed. More specifically, we selected cognitive tasks so that they would successively demand additional sub-components (e.g. the inspection time task as a measure of perceptual speed, the simple reaction time task as a measure of motor speed and basic perceptual processes, and the modified d2 as a measure of mental operation speed beyond perceptual and motor processes). However, these sub-components are not necessarily strictly independent or discrete (Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015; Stafford & Gurney, 2011; Townsend, 1990).

We therefore might have not fully disentangled the specific influences of the different sub-components.

Despite these weaknesses, a major advantage of the present study was the systematic application of different tasks and the experimental manipulation of important task features that have been postulated to influence performance in sustained attention tests. Using an experimental validation approach, we were able to examine different potential causes of the large practice effect in sustained attention tests and altogether, to gain a deeper understanding of the mechanisms that underly performance in these tasks (see also Bornstein, 2011; Borsboom, Mellenbergh, & Van Heerden, 2004).

## Conclusion

The present study set out to explore the locus of the large practice effect in sustained attention tests. Our study showed that it is most likely the item solving process and possibly also the coordination of *perceptual*, *mental operation* and *motor* processes that benefits most from practice, while the practice effects are substantially smaller for *perceptual* and *motor* components. As there are many different sustained attention tests that demand different mental operations (e.g. letter cancellation, mental arithmetics, sorting according to categories), it would be interesting to systematically investigate whether some of these item-solving processes are less susceptible to practice than others and why that is the case. Moreover, the postulated *item shifting* and *preprocessing mechanisms* did not show practice gains. Generally, it has again been shown that the field of psychological assessment greatly profits from experimental methods because they allow a closer look into phenomena like the large practice effects in cognitive ability tests, which is imperative to address or tackle them. We hope that the present paper has taken a first step towards a deeper understanding of practice effects in sustained attention tests.

## References

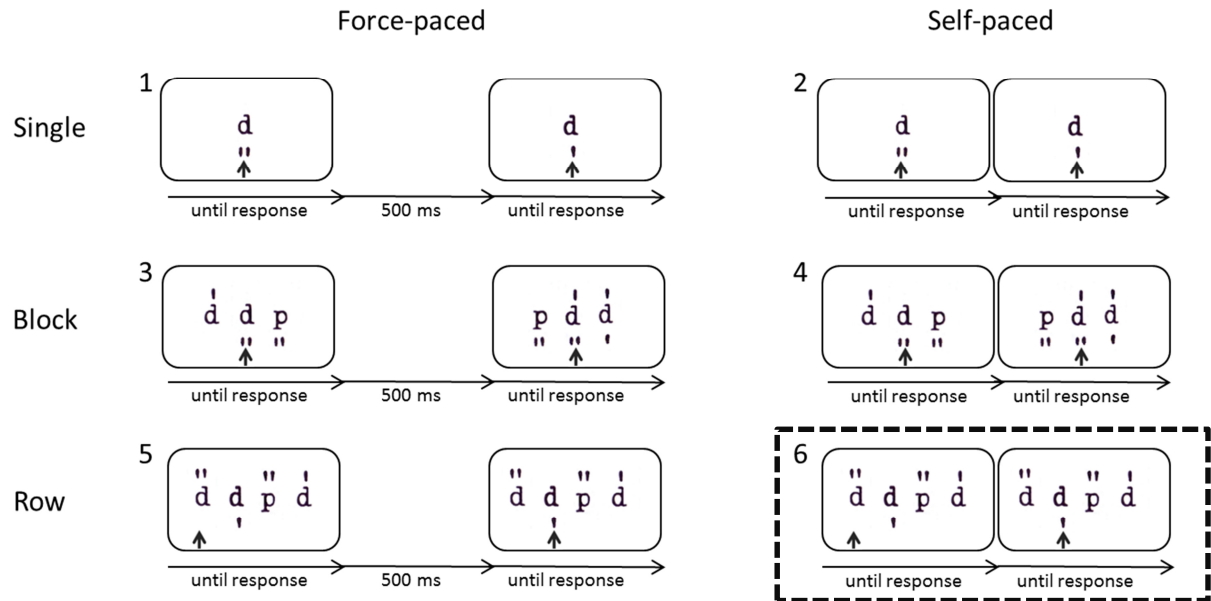
- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117, 288–318.
- Blotenberg, I., & Schmidt-Atzert, L. (2019a). *On the characteristics of sustained attention test performance - The role of the preview benefit*. Manuscript Submitted for Publication.
- Blotenberg, I., & Schmidt-Atzert, L. (2019b). Towards a process model of sustained attention tests. *Journal of Intelligence*, 7, 3.
- Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, 23, 532–544.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.
- Brand, C. R., & Deary, I. J. (1982). Intelligence and “inspection time.” In H. Eysenck (Ed.), *A model of intelligence* (pp. 133–148). New York: Springer.
- Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *Test d2 - Revision. Aufmerksamkeits- und Konzentrationstest (d2-R)*. Göttingen: Hogrefe.
- Bühner, M., Ziegler, M., Bohnes, B., & Lauterbach, K. (2006). Übungseffekte in den TAP Untertests Test Go/Nogo und Geteilte Aufmerksamkeit sowie dem Aufmerksamkeits-Belastungstest (d2). *Zeitschrift für Neuropsychologie*, 17, 191–199.
- Büttner, G., & Schmidt-Atzert, L. (2004). *Diagnostik von Konzentration und Aufmerksamkeit*. Göttingen: Hogrefe.
- Düker, H. (1957). *Leistungsfähigkeit und Keimdrüsenhormone*. München: J.A. Barth.
- Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, 2, 101–118.
- Fleishman, E. A., & Hempel, W. E. (1954). Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 19, 239–252.
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference

- control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133, 101–135.
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373–385.
- Krumm, S., Schmidt-Atzert, L., Bracht, M., & Ochs, L. (2011). Coordination as a crucial component of performance on a sustained attention test. *Journal of Individual Differences*, 32, 117–128.
- Krumm, S., Schmidt-Atzert, L., & Eschert, S. (2008). Investigating the structure of attention: How do test characteristics of paper-pencil sustained attention tests influence their relationship with other attention tests? *European Journal of Psychological Assessment*, 24, 108–116.
- Krumm, S., Schmidt-Atzert, L., Michalczyk, K., & Danthiir, V. (2008). Speeded paper-pencil sustained attention and mental speed tests. *Journal of Individual Differences*, 29, 205–216.
- Krumm, S., Schmidt-Atzert, L., Schmidt, S., Zenses, E. M., & Stenzel, N. (2012). Attention tests in different stimulus presentation modes. *Journal of Individual Differences*, 33, 146–159.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, 58, 981–1007.
- Lievens, F., Reeve, C. L., & Heggstad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology*, 92, 1672–1682.
- Mirsky, A. F., Anthony, B. J., Duncan, C. C., Ahearn, M. B., & Kellam, S. G. (1991). Analysis of the elements of attention: A neuropsychological approach. *Neuropsychology Review*, 2, 109–145.
- Neubauer, A. C., & Knorr, E. (1997). Elementary cognitive processes in choice reaction time tasks and their correlations with intelligence. *Personality and Individual Differences*, 23, 715–728.
- Randall, J. G., & Villado, A. J. (2017). Take two: Sources and deterrents of score change in

- employment retesting. *Human Resource Management Review*, 27, 536–553.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Scharfen, J., Blum, D., & Holling, H. (2018). Response time reduction due to retesting in mental speed tests: A meta-analysis. *Journal of Intelligence*, 6, 6.
- Scharfen, J., Peters, J. M., & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, 67, 44–66.
- Schmidt-Atzert, L., & Brickenkamp, R. (2017). *Test d2-R - Elektronische Fassung des Aufmerksamkeits- und Konzentrationstests d2-R*. Göttingen: Hogrefe.
- Schmidt-Atzert, L., & Bühner, M. (2000). Aufmerksamkeit und Intelligenz. In K. Schweizer (Ed.), *Intelligenz und Kognition: Die kognitiv biologische Perspektive der Intelligenz* (pp. 125–151). Landau: Verlag Empirische Pädagogik.
- Schmidt-Atzert, L., Bühner, M., & Enders, P. (2006). Messen Konzentrationstests Konzentration? Eine Analyse der Komponenten von Konzentrationsleistungen. *Diagnostica*, 52, 33–44.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime 2.0 software. Pittsburgh: Psychology Software Tools Inc.
- Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, and Psychophysics*, 74, 5–35.
- Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A., & Bergmann, K. (2015). Decomposing the relationship between mental speed and mental abilities. *Intelligence*, 51, 28–46.
- Schweizer, K., & Koch, W. (2003). Perceptual processes and cognitive ability. *Intelligence*, 31, 211–235.
- Schweizer, K., & Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence*, 32, 329–347.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15, 72–101.

- Stafford, T., & Gurney, K. N. (2011). Additive factors do not imply discrete processing stages: A worked example using models of the Stroop task. *Frontiers in Psychology*, 2.
- Steinborn, M. B., Langner, R., Flehmig, H. C., & Huestegge, L. (2017). Methodology of performance scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines. *Psychological Assessment*, Advance online publication.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information Processing in Motor Control and Learning* (pp. 117–152). New York: Academic Press.
- Te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, 35, 283–300.
- Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, 1, 46–54.
- Van Breukelen, G. J. P., Roskam, E. E. C. I., Eling, P. A. T. M., Jansen, R. W. T. L., Souren, D. A. P. B., & Ickenroth, J. G. M. (1995). A model and diagnostic measures for response time series on tests of concentration: Historical background, conceptual framework, and some applications. *Brain and Cognition*, 27, 147–179.
- Vickers, D., Nettelbeck, T., & Willson, R. J. (1972). Perceptual indices of performance: The measurement of ‘inspection time’ and ‘noise’ in the visual system. *Perception*, 1, 263–295.
- Vickers, D., & Smith, P. L. (1986). The rationale for the inspection time index. *Personality and Individual Differences*, 7, 609–623.
- Westhoff, K., & Dewald, D. (1990). Effekte der Übung in der Bearbeitung von Konzentrationstests. *Diagnostica*, 36, 1–15.
- Westhoff, K., & Graubner, J. (2003). Konstruktion eines komplexen Konzentrationstests. *Diagnostica*, 49, 110–119.
- Westhoff, K., & Hagemeister, C. (2005). *Konzentrationsdiagnostik*. Lengerich: Pabst.

## Tables and Figures



*Figure 1.* Illustration of the six different conditions of the modified d2, varying stimulus arrangement (single, block, row) and pace (force-paced [with RSI] vs. self-paced [without RSI]). For all of these conditions, the task was to decide whether the relevant stimulus, as indicated by an arrow, was a d2 or not and to press the corresponding key. The bottom right box highlights the condition that closely resembled the original d2-R (“model d2”). In this condition, like in the original test, rows of stimuli were presented simultaneously and required the test-taker to deliberately shift between items in a self-paced mode. Reprinted from “On the characteristics of sustained attention test performance - The role of the preview benefit” by I. Blotenberg and L. Schmidt-Atzert, 2019a, submitted for publication. Reprinted with permission.



Table 1.

*Mean reaction times / inspection time (in ms), standard deviations, retest reliabilities and practice effects (in ms) of the model d2 and the tasks assessing the sub-components of sustained attention test performance.*

Tasks	1 <sup>st</sup> Admini- stration	2 <sup>nd</sup> Admini- stration	Reliability	Practice effect	
	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>r<sub>tt</sub></i>	<i>M</i> ( <i>SD</i> )	<i>Glass' Δ</i>
Model d2	655.18 (117.47)	582.23 (111.33)	.87	72.95 (58.21)	0.62
Sub-Components					
Inspection Time Task (perceptual speed)	53.73 (27.76)	47.35 (31.83)	.58	6.38 (26.00)	0.23
Simple Reaction Time Task (motor speed)	239.60 (26.27)	227.43 (26.15)	.57	12.17 (24.30)	0.46
Simple modified d2 (mental operation speed)	634.00 (95.54)	536.95 (72.52)	.68	97.05 (61.28)	1.02

Table 2.

*Split-half reliability of the practice effects in the subcomponents (assessed as the residuals that capture the change from the first to the second test administration) and correlations (corrected correlations in parentheses) of the practice effect in the model d2 and the sub-components of sustained attention tests.*

	$r_{tt}$	$\Delta$ Inspection Time	$\Delta$ Simple Reaction Time	$\Delta$ Simple modified d2
$\Delta$ Model d2	.47	.14 (.20 <sup>a</sup> )	.04 (.08 <sup>b</sup> )	.33 <sup>**</sup> (.50 <sup>b</sup> )
Sub-Components				
$\Delta$ Inspection Time	. <sup>2</sup>		.21 <sup>*</sup> (.27 <sup>a</sup> )	.23 <sup>*</sup> (.24 <sup>a</sup> )
$\Delta$ Simple Reaction Time	.58			.08 (0.11 <sup>b</sup> )
$\Delta$ Simple modified d2	.91			

Notes: <sup>\*</sup>  $p < .05$ , <sup>\*\*</sup>  $p < .01$ . <sup>a</sup> single correction for attenuation, <sup>b</sup> double correction for attenuation.

<sup>2</sup> Regarding the inspection time task, since it is an adaptive measure, it was not possible to calculate the internal consistency of the residual.

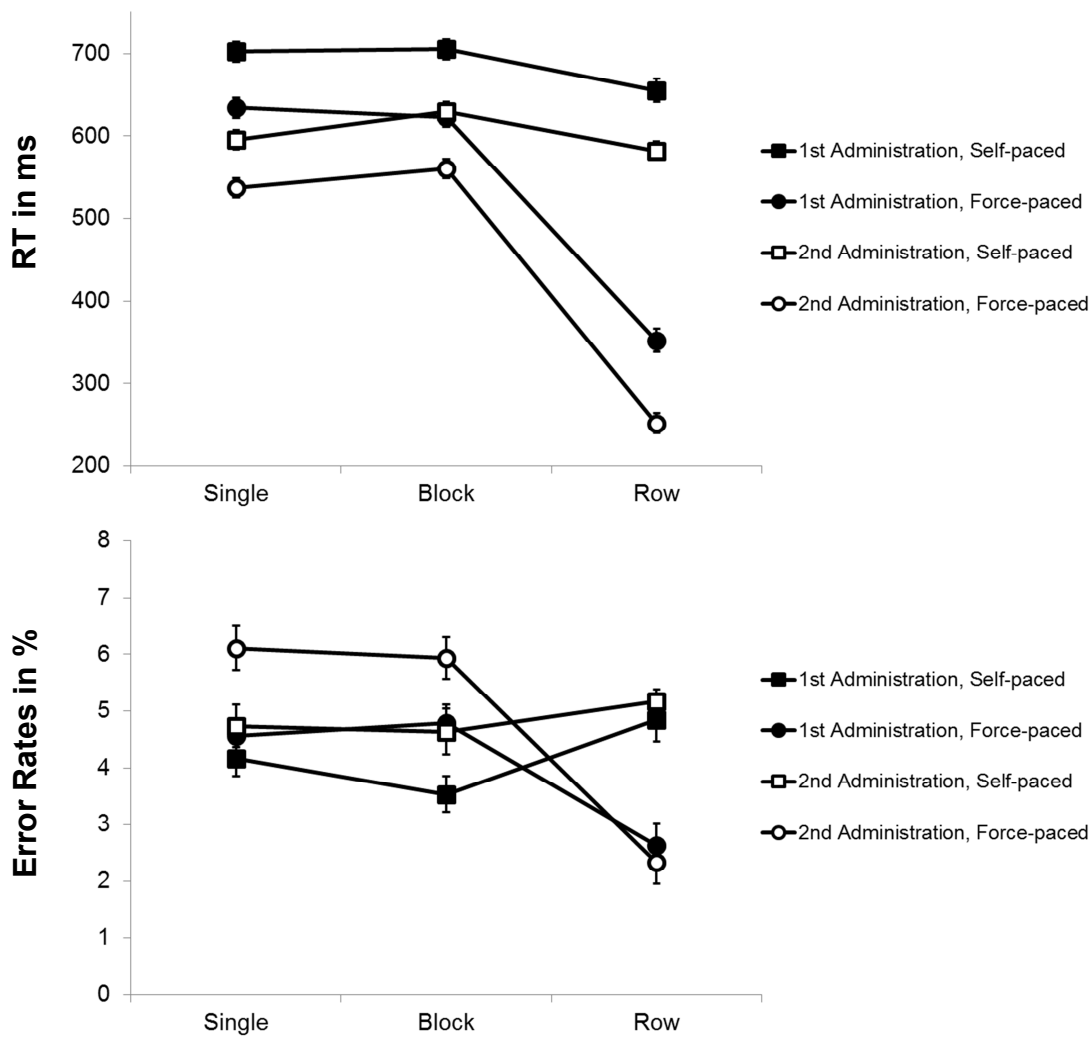


Figure 2. Mean RT (in ms) and error rates (in %) of the modified d2 as a function of stimulus arrangement (single vs. blocks vs. rows of stimuli) and pace (self-paced vs. force-paced) for the first and second administration. Bars represent standard errors.

## Appendix A

Table A1.

*Performance in the original d2-R test of sustained attention regressed on the tasks representing the sub-components of sustained attention tests (stepwise multiple regression).*

Predictors	1 <sup>st</sup> Administration				2 <sup>nd</sup> Administration			
	$\beta$	R <sup>2</sup>	adj. R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	adj. R <sup>2</sup>	$\Delta R^2$
Model 1								
Inspection Time Task (perceptual speed)	.26*	.07*	.06		.31**	.09**	.09	
Model 2								
Inspection Time Task (perceptual speed)	.23*				.26*			
Simple Reaction Time Task (motor speed)	.26*	.13**	.11	.07*	.18*	.13**	.11	.03*
Model 3								
Inspection Time Task (perceptual speed)	.08				.09			
Simple Reaction Time Task (motor speed)	.02				.02			
Simple Modified d2 (mental operation speed)	.67**	.47**	.46	.34**	.60**	.41**	.40	.29**

Notes: \*  $p < .05$ , \*\*  $p < .01$ .

**Eidesstattliche Erklärung**

Hiermit erkläre ich, dass ich meine Dissertation

*„Über die Charakteristika der Leistung in Konzentrationstests – Prozesskomponenten, Mechanismen und Übungseffekte“*

selbständig und ohne unerlaubte Hilfe angefertigt habe und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, im Februar 2019

Iris Blotenberg